Министерство образования и науки Российской Федерации

ФГБОУ ВО "Тамбовский государственный технический университет"

КАРПУШКИН С.В., ГЛЕБОВ А.О.

ТЕОРИЯ ИНЖЕНЕРНОГО ЭКСПЕРИМЕНТА

Учебное пособие для студентов дневного и заочного отделения,
обучающихся по направлениям
15.03.01 "Машиностроение", 15.04.01 "Машиностроение",
15.04.05 "Конструкторско-технологическое обеспечение машиностроительных производств"

Тамбов
2017

Рецензенты:

Громов М.С. – к.т.н., технический директор
ООО «СпиртПромПроект», г. Тамбов

Борщев В.Я. – д.т.н., профессор кафедры «Технологические процессы, аппараты и техносферная безопасность», ТГТУ

Утверждено Методическим советом ТГТУ
(протокол № 3 от 22.06.2017 г.)

ОГЛАВЛЕНИЕ

ВВЕДЕНИЕ

В настоящем учебном пособии рассматриваются практические рекомендации по построению математических моделей объектов исследования на основе результатов экспериментов в виде уравнений регрессии. Излагается методика планирования эксперимента – определения числа и условий проведения опытов, обеспечивающих требуемый уровень достоверности формируемых математических моделей.

В пособии даны рекомендации по выбору факторов эксперимента, откликов объекта исследования, вида уравнения регрессии, рассмотрены основные понятия математической статистики и стохастической связи между случайными величинами. Основное внимание уделено методике предварительной обработки результатов экспериментов, формирования и проверке адекватности уравнения регрессии объекту исследования. Рассмотрена методика проведения и обработки результатов полного и дробного факторного эксперимента, разновидности центральных композиционных планов эксперимента.

Пособие содержит лабораторный практикум, включающий задания и методические указания по выполнению четырех лабораторных работ, выполнение которых способствует выработке практических навыков постановки экспериментов, предварительной обработки их результатов, формированию и проверке адекватности уравнений регрессии, реализации операций их интерпретации.

Приложения к пособию содержат табличные значения критериев Стьюдента, Кохрена и Фишера при уровне значимости 0.05, необходимые для предварительной обработки результатов экспериментов, проверки адекватности уравнений регрессии.

1 ЭКСПЕРИМЕНТ И ОБРАБОТКА ЭКСПЕРИМЕНТАЛЬНЫХ ДАННЫХ ДЛЯ КОНКРЕТНОГО ОБЪЕКТА ИССЛЕДОВАНИЯ

Рассмотрим конкретный пример постановки и решения инженерной исследовательской задачи. Пусть объектом исследования является котел, производящий тепло в количестве Q Дж, затрачивая топливо в количестве V м^3/с, если топливо – газ, или M кг/с, если топливо – мазут. Задача оптимизации работы котла: отношение Q/V или Q/M должно быть максимальным, насколько это возможно.

Одним из подходов к решению подобных задач является разработка математической модели функционирования объекта исследования: совокупности соотношений, определяющих влияние параметров объекта и условий его функционирования $x_1, x_2, ..., x_n$ на результат его работы y. Чаще всего в инженерной практике стремятся сформировать функцию типа $y = f(x_1, x_2, ..., x_n)$, с помощью которой можно прогнозировать поведение объекта при изменении его параметров и условий работы, а также определять значения $x_1, x_2, ..., x_n$, соответствующие экстремальным значениям y. Популярным способом формирования подобных функций является использование методики планирования эксперимента.

1.1 Планирование эксперимента

Прежде всего определяется круг исследуемых параметров объекта и условий его функционирования, называемых факторами $x_1, x_2, ..., x_n$. Для нашего примера выберем четыре фактора:

x_1 – давление пара,

x_2 – его расход,

x_3 – площадь поверхности теплообмена между продуктами сгорания топлива и паром,

x_4 – вид топлива.

Каждый фактор имеет допустимый диапазон значений, например, давление пара не должно превосходить давления его насыщения при данной температуре.

Эксперимент заключается в фиксировании изменения теплоотдачи котла, вызванного изменениями значений факторов. Он должен быть спланирован так, чтобы получить максимальное количество информации при минимальных затратах средств и времени. При планировании эксперимента внутри диапазона допустимых изменений значения каждого фактора (от $x_{i\,min}$ до $x_{i\,max}$, $i = 1,2,3,4$) необходимо выбрать ряд промежуточных значений, например, девять возможных значений (от $x_{i1} = x_{i\,min}$ до $x_{i9} = x_{i\,max}$). Для нашего примера запланируем 50 опытов, при этом сочетания значений факторов для каждого опыта зададим случайным образом и получим так называемую таблицу планирования эксперимента:

№ п/п	x_1	x_2	x_3	x_4	y
1	x_{18}	x_{22}	x_{33}	x_{47}	
2	x_{11}	x_{29}	x_{34}	x_{44}	
...	...	...	...	...	
50	x_{17}	x_{23}	x_{36}	x_{45}	

Далее назначается период времени экспериментальной работы котла по режиму каждой строки таблицы, например, одна неделя. В течение этого периода фиксируется теплоотдача котла и ее показатели, усредненные за этот период, заносятся в таблицу в виде экспериментального значения величины $y = Q/V$ или Q/M: y_i, $i = 1,...,50$.

Когда таблица будет заполнена полностью, все ее содержимое составит так называемые *экспериментальные данные*, а сама таблица будет называться *таблицей экспериментальных данных* и представлять собой таблично заданную функцию – зависимость эффективности работы котла y от изменения значений факторов x_1, x_2, x_3, x_4.

Цель обработки экспериментальных данных заключается в том, чтобы эту табличную, аналитически неизвестную зависимость между переменной y и переменными x_1, x_2, x_3, x_4 представить в виде математической модели, т.е. уравнения, которое "достаточно точно" согласовывало бы расчётные и табличные значения отклика объекта y, например, $y = \alpha_1 x_1 + \alpha_2 x_2 + \alpha_3 x_3 + \alpha_4 x_4$.

1.2 Основные термины и положения

Объект исследования – это объект любого характера (технического, социального, экономического и т.д.), который изучается экспериментально.

Факторы **–** это элементы воздействия на объект.

Отклики объекта исследования – это его реакции на воздействия.

Эксперимент – это специальным образом спланированная и организованная *процедура изучения* некоторого объекта исследования, в ходе которой на него осуществляются запланированные воздействия и регистрируются его отклики на них. Эксперимент состоит из ряда *опытов* (или *наблюдений*), которым соответствуют разные значения каждого из факторов.

Экспериментальные данные – все исходные и выходные числовые данные эксперимента, сведенные в *таблицу экспериментальных данных*.

Обработка экспериментальных данных – различные методы построения математической модели объекта по таблице экспериментальных данных.

Регрессионный анализ – наиболее распространенный метод обработки данных экспериментов, который предусматривает получение зависимостей вида $y = f(x_1,x_2,..., x_n)$, соответствующих таблице экспериментальных данных (эти зависимости называют *уравнениями регрессии*).

Управляемые факторы – это такие воздействия на объект исследования, численные значения которых определяются и контролируются самим экспериментатором.

Активный эксперимент – это эксперимент, в котором задействованы только управляемые факторы. Пример – изучение зависимости выхода целевого продукта реакции от температуры в реакторе.

Контролируемые факторы – это такие воздействия на объект исследования, численные значения которых экспериментатором не устанавливаются, но их значения исследователь может контролировать и фиксировать.

Пассивный эксперимент – это эксперимент, в котором задействованы только контролируемые факторы. Пример – изучение зависимости выхода целевого продукта реакции от химического состава используемого сырья.

Активно-пассивный (или пассивно-активный) эксперимент – это совмещение обоих видов эксперимента, например, изучение зависимости выхода целевого продукта реакции от температуры в реакторе и химического состава сырья.

Основным "рабочим инструментом" и эксперимента и обработки экспериментальных данных являются численные значения факторов воздействия и откликов объекта исследования, т.е. *числа*. Какова ни была бы природа факторов и откликов, включая, в том числе, эмоции или впечатления, они должны быть выражены количественно, числом.

Числовые значения факторов и откликов эксперимента можно получить *подсчетом, измерением и методом экспертных оценок***.** Примером последнего способа является оценка, выставляемая студенту преподавателем на экзамене.

2 ТОЧНОСТЬ И ПОГРЕШНОСТИ ВЫЧИСЛЕНИЙ, СПОСОБЫ ИХ ОЦЕНКИ И УМЕНЬШЕНИЯ ПОГРЕШНОСТЕЙ

Анализ точности результата вычислений, т.е. определение его последней значащей цифры, которой можно доверять, является важной составной частью вычислительного процесса. Реальная оценка практически достижимого уровня точности способствует не только экономии сил и средств, но часто связана с вопросами достоверности достигнутых результатов и безопасности их прикладного использования.

2.1 Понятие приближенного числа и погрешности

Анализ точности результата вычислений осуществляется на основе понятия ***погрешности***. Пусть x_1 – истинное значение некоторой величины, а x_{1r} – значение, присвоенное ей в ходе эксперимента или экспертной оценки, при которых всегда возникают некоторые ошибки, например, результат любого измерения содержит ошибку датчика и прибора. Поэтому значение x_{1r} называют *приближенным* значением изучаемой величины или просто *приближенным* числом.

Абсолютная погрешность Δx любого приближенного числа есть абсолютная величина разности между истинным значением величины x_1 и ее приближенным значением x_{1r}, т.е. $\Delta x = |x_1 - x_{1r}|$. Истинное значение величины, как правило, неизвестно, поэтому оценкой абсолютной погрешности считается установление неравенства вида $|x_1 - x_{1r}| \le \Delta x_p$, где Δx_p – предельная абсолютная погрешность.

Абсолютная погрешность записи приближенных чисел, обычно принимается равной половине единицы последнего разряда в записи, например, запись числа π = 3.14 имеет погрешность 0.005, а π = 3.1416 – 0.00005. Если обозначение числа имеет большее количество цифр, чем это требуется согласно установленной точности, их следует округлить.

В окончательных результатах расчета принято записывать числа, сохраняя одну недостоверную цифру за последней верной, а предельную абсолютную погрешность указывают за этим результатом после знака "±". Например, если при расчете получено число 271.734 с предельной абсолютной погрешностью 0.043, то последняя достоверная цифра после запятой «7», а последняя записываемая цифра «3» и, согласно правилам, результат должен быть записан как 271.73 ± 0.05.

Абсолютная погрешность характеризует точность приближенного числа недостаточно: погрешность в 0.5 метра слишком велика для отмеривания куска ткани в магазине и слишком мала для измерения расстояния между двумя городами. Таким образом, пригодность абсолютной погрешности выявляется только при сопоставлении с практическим значением данной переменной, т.е. по так называемой ***относительной*** погрешности. Относительная погрешность dx определяется как отношение абсолютной погрешности Δx (или Δx_p) к абсолютной величине истинного значения x_1, т.е. $dx = \Delta x_p / |x_{1r}|$. Иногда это отношение умножают на 100, т.е. выражают в процентах (процентная погрешность). На практике обычно пользуются понятием предельной относительной погрешности dx_p, которая удовлетворяет неравенству $dx \le dx_p$.

Погрешности в силу разных источников их происхождения классифицируют как ***устранимые*** (*инструментальные*, *методические)* и ***неустранимые*** (*наследственные*).

Инструментальные погрешности связаны с конечной точностью представления исходной информации (округление значений входных величин или ошибки измерений).

Методические погрешности обусловлены тем, что многие задачи решаются приближенно с использованием специальных численных методов. Это, в частности, относится к вычислению тригонометрических, логарифмических, показательных функций.

Наследственные погрешности – это погрешности результата вычислений, вызванные распространением или трансформацией погрешностей исходных данных при прохождении их по вычислительному алгоритму через ряд промежуточных результатов.

2.2 Оценка погрешностей вычислительного процесса

Пусть независимая переменная x известна с некоторой точностью. С какой точностью можно найти значение функции $y = f(x)$? Обратная задача: если необходимо рассчитать значение функции y с заданной точностью, то какова должна быть точность определения значения переменной x?

В большинстве технических расчетов удовлетворительным считается такой уровень точности результата, при котором его максимальная относительная погрешность не превышает 5 % . Например, при переходе системы единиц МКГСС к Международной системе единиц (СИ) перевод килограмма силы в Ньютоны можно осуществить введением множителя 10 вместо более точного множителя 9,807. Ошибка при этом составит (10 – 9.81)/9.81 = 0.02, что в большинстве случаев вполне допустимо.

Однако, такой уровень точности бывает и неприемлемым. Например, при запуске искусственных спутников Луны с околоземной орбиты вторую космическую скорость (около 11200 м/с) требуется выдержать с ошибкой, не превышающей 0.000002. В противном случае запускаемый аппарат станет спутником не Луны, а Солнца.

Трансформация наследственных погрешностей при осуществлении вычислительного процесса осуществляется по определенным закономерностям. Наиболее важные случаи распространения ошибок:

Операция	Реализация операции	Абсолютная ошибка	Относительная ошибка
Сложение	$x_1 + x_2$	$\Delta x_1 + \Delta x_2$	$(\Delta x_1 + \Delta x_2)/\lvert x_1 + x_2\rvert$
Вычитание	$x_1 - x_2$	$\Delta x_1 + \Delta x_2$	$(\Delta x_1 + \Delta x_2)/\lvert x_1 - x_2\rvert$
Умножение	$x_1 \times x_2$	$\Delta x_1 \times \lvert x_2\rvert + \Delta x_2 \times \lvert x_1\rvert$	$\Delta x_1/\lvert x_1\rvert + \Delta x_2/\lvert x_2\rvert$
Деление	x_1 / x_2	$(\Delta x_1 \times \lvert x_2\rvert + \Delta x_2 \times \lvert x_1\rvert)/x_2^2$	$\Delta x_1/\lvert x_1\rvert + \Delta x_2/\lvert x_2\rvert$
Возведение в степень*	x^n	$\Delta x \times \lvert n\rvert \times \lvert x^{n-1}\rvert$	$\Delta x \times \lvert n\rvert/\lvert x\rvert$

*показатель степени n может принимать произвольные значения, если n – правильная дробь, то ошибка уменьшается.

Пример 1. Диаметр круга D определен с некоторой погрешностью. Как определить погрешность величины площади круга S, вычисленной по формуле $S = \pi \cdot D^2/4$?

Измеренный диаметр D составил 5 см, абсолютная погрешность для данных условий изготовления круга и измерения его диаметра оценивается как ±2 мм. Тогда предельная относительная погрешность равна 0.2/5 = 0.04 и поскольку в вычислении заложена операция умножения $D \times D$, то относительная погрешность вычисленной площади S круга составит

$$\Delta S/S = \Delta D/D + \Delta D/D = 0.04 + 0.04 = 0.08.$$

Тогда абсолютная погрешность площади круга

$$\Delta S = 0.08 \times S = 0.08 \times 19.625 = 1.57 \text{ см}^2.$$

Пример 2. Известно, что площадь квадрата равна 12.34 см^2.С какой погрешностью должна быть измерена сторона квадрата, чтобы расчет его площади обеспечил необходимую точность?

Из выражения площади следует, что она дана с предельной абсолютной погрешностью 0.005 см^2, тогда относительная погрешность составит (0.005/12.34) = 0.0004. Так как длина стороны квадрата L есть $S^{0.5}$, то $\Delta L/L = |0.5| \times 0.04 = 0{,}0002$, откуда абсолютная погрешность измерения стороны квадрата составит $\Delta L = S \times 0.0002 = 0.0007$ см.

Способы уменьшения значения наследственных погрешностей:

- при выполнении операций сложения и вычитания над последовательностью чисел, содержащей большое количество членов, следует сначала оперировать с числами, наименьшими по абсолютной величине;
- следует избегать вычитания близких по значению чисел;
- следует избегать сложения чисел, отличающихся на несколько порядков;
- для уменьшения погрешностей округления чисел промежуточные действия рекомендуется производить, сохраняя после запятой на 1-2 знака больше, чем требуется в окончательном результате.

3 МАТЕМАТИЧЕСКАЯ МОДЕЛЬ ОБЪЕКТА ИССЛЕДОВАНИЯ В ВИДЕ АЛГЕБРАИЧЕСКОГО ПОЛИНОМА

3.1 Основные задачи исследования и назначение математической модели

Термин "исследование" включает предварительную подготовку эксперимента (сбор, анализ и обработку исходных данных), проведение эксперимента и обработку полученных результатов. К числу основных задач исследования относятся:

1) статистический анализ (точечное и интервальное оценивание, проверка справедливости статистических гипотез) вовлекаемых в эксперимент факторов с целью выделения факторов, существенно влияющих на отклик объекта исследования. При практическом планировании эксперимента решается задача оптимизации – определение комбинации значений (уровней) управляемых факторов, которой соответствует экстремум функции отклика;

2) идентификация функции отклика объекта $\varphi(\boldsymbol{x})$, где $\boldsymbol{x} = (x_1,...,x_i,...,x_m)$ – вектор факторов, вовлекаемых в эксперимент: установление тождественности $\varphi(\boldsymbol{x})$ алгебраическому степенному полиному $\eta(\boldsymbol{x},\boldsymbol{\beta})$, $\boldsymbol{\beta} = (\beta_0,..., \beta_j,..., \beta_n)$ –вектор коэффициентов функции $\eta(\boldsymbol{x},\boldsymbol{\beta})$, которая называется идеальной математической моделью функции отклика $\varphi(\boldsymbol{x})$. Такие модели могут быть линейно или нелинейно параметризованными по коэффициентам $\boldsymbol{\beta}$.

На практике чаще всего используются линейные функции $\eta(\boldsymbol{x},\boldsymbol{\beta})$:

$$\eta(\boldsymbol{x},\boldsymbol{\beta}) = \beta_0 + \beta_1 \cdot x_1 + ... + \beta_m \cdot x_m,$$

а в общем случае

$$\eta(\boldsymbol{x},\boldsymbol{\beta}) = \beta_0 + \sum_{j=1}^{n} \beta_j \cdot f_j(\boldsymbol{x}),$$

где $f_j(\boldsymbol{x})$ – сумма степенных функций $x_1,...,x_m$, которая может содержать произведения различных факторов ($x_i \cdot x_k$, $i \neq k$; $x_i \cdot x_k \cdot x_l$, $i \neq k \neq l$).

Получение математической модели рассматриваемого объекта позволяет:

- выделить значимые факторы, изменения значений которых существенно влияют на значение отклика объекта, а также исследовать зависимости соотношений значений факторов (т.е. условий функционирования объекта) и значениями его откликов;
- предсказать значение отклика для заданных комбинаций уровней факторов, решая задачи как интерполяции, так и экстраполяции (последнее – в определенных пределах ;
- находить координаты точек минимумов и максимумов функции $\eta(\boldsymbol{x},\boldsymbol{\beta})$;
- уточнять гипотетические и теоретические положения, высказываемые относительно объекта исследования;
- выдвигать новые гипотезы о процессах, протекающих в объекте.

3.2 Алгебраический полином как математическая модель объекта исследования

Метод регрессионного анализа использует описание объекта исследования в виде такой формы алгебраического полинома, которая в качестве функций $f_j(\boldsymbol{x})$ содержит все возможные произведения факторов $x_1,x_2,...,x_m$ в первой степени (единичные, парные, тройные и т.д.), а при степени больше единицы – только их единичные индивидуальные комбинации:

$$\varphi(x_1,x_2,...,x_m)=\beta_0+\sum_{i=1}^{m}\beta_i\cdot x_i+\sum_{i=1}^{m}\beta_{i,i+1}\cdot x_i\cdot x_{i+1}+...+\sum_{i=1}^{m}\beta_{i,i+1,...,i+m}\cdot x_i\cdot x_{i+1}\cdot...\cdot x_m+\sum_{i=1}^{m}\beta_{ii}\cdot x_i^2$$

$+\sum_{i=1}^{m}\beta_{iii}\cdot x_i^3+$ ..., где коэффициенты $\boldsymbol{\beta}$ необходимо подобрать так, чтобы значения функции $\varphi(x_1,x_2,...,x_m)$ соответствовали отклику объекта исследования.

Найти полином, обеспечивающий получение указанного результата, как правило, не удается, т.к. состояние любого реального объекта определяется очень большим количеством факторов, и любая модель объекта принципиально не может быть полной, а только приближенной. Кроме того, точный вид полинома, адекватный функции $\varphi(\boldsymbol{x})$, не известен так же, как и сама эта функция.

Поэтому та зависимость, которая будет найдена по таблице экспериментальных данных, не дает точного представления о связи между откликом объекта и факторами, включенными в математическую модель, т.е. по результатам эксперимента находится только статистическая оценка этой связи в виде уравнения

$$y(x_1,x_2,...,x_m)=b_0+\sum_{i=1}^{k_1}b_i\cdot x_i+\sum_{i=1}^{k_2}b_{i,i+1}\cdot x_i\cdot x_{i+1}+...\sum_{i=1}^{k_{m+2}}b_{iii}\cdot x_i^3+...,\qquad(3.1)$$

где $\boldsymbol{b}$ – "выборочные" эмпирические коэффициенты регрессии, которые являются оценками теоретических коэффициентов $\boldsymbol{\beta}$.

Практика обработки экспериментальных данных показала, что результаты эксперимента в виде табличной функции в большинстве случаев с достаточной точностью приближаются полным кубическим полиномом в форме (3.1). Часто третья степень полинома не только достаточна, но и избыточна, т.е. степень и количество членов полинома можно и уменьшить без существенной потери точности. Поэтому при построении и выборе аппроксимирующего уравнения строят ***систему альтернативных уравнений***, которая включает полный кубический полином и полиномы, образованные из полного путем присвоения некоторым из его коэффициентов нулевых значений.

3.3 Альтернативные уравнения регрессии

Сравнивая характеристики степени соответствия этих уравнений таблице экспериментальных данных, выбирают наиболее приемлемое, например, с точки зрения минимума среднеквадратичного отклонения расчетных данных от экспериментальных. В качестве примера такого подхода рассмотрим кубическое уравнение для 5-ти факторной задачи регрессии. Полный кубический пятифакторный полином имеет вид:

$$y=b_0+b_1x_1+...+b_5x_5+b_{12}x_1x_2+...+b_{15}x_1x_5+b_{23}x_2x_3+...+b_{25}x_2x_5+...$$
$$+b_{35}x_3x_5+b_{45}x_4x_5+b_{123}x_1x_2x_3+...+b_{125}x_1x_2x_5+...+b_{135}x_1x_3x_5+...$$
$$+b_{245}x_2x_4x_5+...+b_{345}x_3x_4x_5+b_{1234}x_1x_2x_3x_4+b_{1235}x_1x_2x_3x_5+$$
$$+b_{1345}x_1x_3x_4x_5+b_{2345}x_2x_3x_4x_5+b_{12345}x_1x_2x_3x_4x_5+b_{11}x_1^2+...$$
$$+b_{55}x_5^2+b_{111}x_1^3+...+b_{555}x_5^3.$$

Этот полином и будет первым альтернативным уравнением регрессии. Далее приведем альтернативные уравнения в виде частей этого полинома различных степеней.

Линейное уравнение:

$$y= b_0+b_1x_1+b_2x_2+b_3x_3+b_4x_4+b_5x_5.$$

Неполное квадратичное уравнение, содержащее линейную часть и парные сочетания факторов:

$$y= b_0+b_1x_1+...+b_5x_5+b_{12}x_1x_2+b_{13}x_1x_3+\ldots+b_{23}x_2x_3+b_{24}x_2x_4+... +b_{35}x_3x_5+b_{45}x_4x_5.$$

Неполное кубическое уравнение, содержащее линейную часть, парные и тройные сочетания факторов:

$$y= b_0+b_1x_1+...+b_5x_5+ b_{12}x_1x_2+...+b_{45}x_4x_5+b_{123}x_1x_2x_3+...+ b_{345}x_3x_4x_5.$$

Неполное уравнение четвертой степени, содержащее линейную часть, парные, тройные и четверные сочетания факторов:

$$y=b_0+b_1x_1+...+b_5x_5+b_{12}x_1x_2+...+b_{45}x_4x_5+b_{123}x_1x_2x_3+...$$
$$+b_{345}x_3x_4x_5+b_{1234}x_1x_2x_3x_4+\ldots+b_{2345}x_2x_3x_4x_5.$$

Следующее уравнение получим по аналогии – добавлением к предыдущему уравнению члена из сочетания "по пять из пяти": $+b_{12345}x_1x_2x_3x_4x_5$.

Седьмое и последнее уравнение – полный полином второй степени:

$$y=b_0+b_1x_1+...+b_5x_5+b_{12}x_1x_2+...+b_{15}x_1x_5+b_{23}x_2x_3+...+b_{25}x_2x_5+...$$
$$+b_{35}x_3x_5+b_{45}x_4x_5+ b_{123}x_1x_2x_3+...+b_{125}x_1x_2x_5+...+b_{135}x_1x_3x_5+...+b_{245}x_2x_4x_5+...$$
$$+b_{345}x_3x_4x_5+b_{1234}x_1x_2x_3x_4+b_{1235}x_1x_2x_3x_5+b_{1345}x_1x_3x_4x_5+b_{2345}x_2x_3x_4x_5+$$
$$+b_{12345}x_1x_2x_3x_4x_5+b_{11}x_1^2+...+b_{55}x_5^2.$$

Таким образом, получена система из семи альтернативных уравнений, в которой обычно удается найти приемлемое уравнение регрессии.

На практике широко используются упрощенные формы записи уравнений регрессии, например, запись только коэффициентов с индексами вида

$$b_0+b_1+\ldots+b_{12}+\ldots+b_{123}+\ldots+b_{1234}+\ldots+b_{12345}+b_{11}+\ldots+b_{111}+\ldots+b_{555}=y,$$

либо запись уравнения только в индексах коэффициентов $\boldsymbol{b}$:

0 1 2 3 4 5 12 13 14 15 23 24 25 34 35 45 123 124 125 134 135 145 234 235 245 345 1234 1235 1245 1345 2345 12345 11 22 33 44 55 111 222 333 444 555.

Замечание. Построение полиномов регрессии именно такой структуры – это *рекомендация*. Одинаково "правомочна" любая другая форма полиномов регрессии.

С увеличением числа факторов, включенных в модель объекта исследования, количество членов полинома быстро нарастает. Так, например, полный кубический полином при трех факторах имеет 14 слагаемых, при четырех – 24, при пяти – 42, а при десяти – уже 1044.

В уравнениях регрессии неизвестными являются значения коэффициентов $\boldsymbol{b}$ (т.к. значения факторов и откликов известны из таблицы экспериментальных данных). Для нахождения каждого коэффициента b необходима одна строка таблицы экспериментальных данных, см. п. 3.4, т.е. при десяти факторах и недельной продолжительности одного опыта построение полного кубического регрессионного полинома займет более 20 лет. Поэтому на практике всегда стремятся максимально укоротить полином регрессии.

Следовательно, по результатам эксперимента будет сформирована ***приближенная*** модель в условиях недостатка информации об изучаемом объекте.

3.4 Полином регрессии и система условных уравнений

Процедура регрессионного анализа начинается с выдвижения гипотезы о конкретном виде уравнения, которым мы намереваемся приблизить экспериментальную зависимость. Вид уравнения регрессии задается либо на основе известных фундаментальных зависимостей, либо перебора нескольких вариантов уравнений и сравнения их по точности воспроизведения табличного значения отклика $\boldsymbol{y}$.

Процедура обработки экспериментальных данных начинается с подстановки в уравнение значений факторов x_{kn} в соответствии со строками таблицы данных, где n – номер строки таблицы, k – номер значения компоненты вектора $\boldsymbol{x}$. В результате будет сформирована система уравнений, каждое из которых соответствует одной строке таблицы экспериментальных данных.

Например, таблицу данных с двумя факторами x_1 и x_2 при числе строк $\boldsymbol{N}$=7 можно приблизить уравнением $b_0 + b_1 \cdot x_1 + b_2 \cdot x_2 + b_{12} \cdot x_1 \cdot x_2 + b_{11} \cdot x_1^2 + b_{22} \cdot x_2^2 = y$. Заметим, что левая часть полинома представляет собой произведение двух векторов: вектора коэффициентов $\boldsymbol{b}$ и вектора множителей при этих коэффициентах $(1 \quad x_1 \quad x_2 \quad x_1 \cdot x_2 \quad x_1^2 \quad x_2^2)$, который называется вектором базисных функций.

Если индексами при коэффициентах $\boldsymbol{b}$ будем обозначать комбинацию базисных функций при данном коэффициенте, а индексами при факторах $\boldsymbol{x}$ – номер фактора и номер строки таблицы, то сформированная система уравнений будет следующей:

$$b_0 + b_1 x_{11} + b_2 x_{21} + b_{12} x_{11} x_{21} + b_{11} x_{11} x_{11} + b_{22} x_{21} x_{21} = y_1;$$
$$b_0 + b_1 x_{12} + b_2 x_{22} + b_{12} x_{12} x_{22} + b_{11} x_{12} x_{12} + b_{22} x_{22} x_{22} = y_2;$$
$$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$$
$$b_0 + b_1 x_{17} + b_2 x_{27} + b_{12} x_{17} x_{27} + b_{11} x_{17} x_{17} + b_{22} x_{27} x_{27} = y_7.$$

Однако, при воздействии на объект исследования факторами $\boldsymbol{x}$, наличие и значение которых определяется экспериментатором, значение отклика $\boldsymbol{y}$ формируется как за счет факторов $\boldsymbol{x}$, так и за счет неучтенных в эксперименте факторов $\boldsymbol{w}$ (т.н. шума).

Если многократно повторять наблюдения, задавая ***одинаковые*** значение факторов $x_{1n}, x_{2n}, \ldots, x_{kn}$ для одной и той же n-ой строки таблицы экспериментальных данных, значения откликов при этом из-за наличия шума будут разными, т.е. значение ошибки наблюдения при повторных опытах не будет одинаковым.

В связи со случайным характером отклика $\boldsymbol{y}$ левая и правая часть полученной выше системы уравнений неравны, система является несовместной, т.е. не существует такой комбинации неизвестных коэффициентов $\boldsymbol{b}$, которая отвечала бы всем уравнениям системы. Поэтому такие системы носят название системы ***условных*** уравнений.

Представим эту систему в другом виде

$$y_1 - (b_0 + b_1x_{11} + b_2x_{21} + b_{12}x_{11}x_{21} + b_{11}x_{11}x_{11} + b_{22}x_{21}x_{21}) = e_1;$$
$$y_2 - (b_0 + b_1x_{12} + b_2x_{22} + b_{12}x_{12}x_{22} + b_{11}x_{12}x_{12} + b_{22}x_{22}x_{22}) = e_2;$$
....
$$y_7 - (b_0 + b_1x_{17} + b_2x_{27} + b_{12}x_{17}x_{27} + b_{11}x_{17}x_{17} + b_{22}x_{27}\,x_{27}) = e_7,$$
где e_n − разность между левой и правой частями исходных уравнений. Эту невязку можно трактовать как отклонения расчетного значения отклика от экспериментального**.** Суммарной характеристикой этих отклонений является ***остаточная сумма*** SUM_{ost}:

$$SUM_{ost} = \sum_{n=1}^{N} (y_n - yr_n)^2 = \sum_{n=1}^{N} e_n^{\,2},$$

где $yr_n = b_0 + b_1x_{1n} + b_2x_{2n} + b_{12}x_{1n}\,x_{2n} + b_{11}x_{1n}\,x_{1n} + b_{22}x_{2n}\,x_{2n}$ – расчетное значение отклика.

Величина SUM_{ost} позволяет сформулировать понятие ***наилучшего*** решения системы условных уравнений. Наилучшим будет решение, которое ***минимизирует*** остаточную сумму. Такой подход называется ***методом наименьших квадратов***. В точке минимума функции SUM_{ost} ее производные $\partial SUM_{ost}/\partial b_j$ равны нулю. Дифференцируя функцию $SUM_{ost}(\boldsymbol{b})$ по всем коэффициентам регрессии и приравнивая производные нулю, получим систему ***нормальных*** уравнений, которая имеет единственное решение и минимизирует остаточную сумму. Подробнее метод наименьших квадратов рассмотрен в разделе 10.2.

Нахождение вектора коэффициентов ***b***, т.е. формирование и решение уравнения регрессии, составляет первую часть процедуры регрессионного анализа. После нахождения полинома регрессии следует оценить его адекватность функции истинного отклика, т.е. точность, с которой уравнение регрессии приближает таблицу экспериментальных данных. Решение этой задачи составляет вторую часть процедуры регрессионного анализа.

4 ВЫБОР ОТКЛИКА ОБЪЕКТА ИССЛЕДОВАНИЯ, ФАКТОРОВ И ВИДА УРАВНЕНИЯ РЕГРЕССИИ

Выбор отклика объекта исследования является одним из главных этапов работы на стадии его предварительного изучения, т.к. правильная постановка задачи исследования зависит от правильности выбора параметров, образующих функцию $y = \varphi(x_1,x_2,...,x_m)$.

4.1 Требования к отклику объекта исследования

Первое: отклик объекта должен быть *количественным*, т.е. задаваться числом. Множество значений, которые может принимать отклик, называется областью его определения, которая может быть непрерывной и дискретной, ограниченной и неограниченной. Например, выход химической реакции – это отклик с непрерывной ограниченной областью определения 0÷100%, а случайное натуральное число – отклик с дискретной неограниченной областью определения.

Если количественная оценка отклика невозможна, используют ранжирование, т.е. присвоение отклику оценки – ранга по заранее выбранной шкале: двухбалльной, пятибалльной и т.д.

Второе: отклик объекта должен выражаться *одним числом* (регистрация показания прибора), либо ему должен сопутствовать алгоритм, применение которого дает одно число, например, вместо множества значений можно использовать среднее арифметическое.

Третье: отклик объекта должен быть *однозначным* в статистическом смысле, т.е. заданному набору значений факторов должно соответствовать одно значение отклика. Заметим, что обратное неверно: одному и тому же значению отклика могут соответствовать разные наборы значений факторов.

Четвертое: отклик объекта должен давать *возможность эффективной оценки функционирования объекта*. Для выполнения этого требования возможно изменение отклика в ходе исследования, например, на первых стадиях исследования технологических процессов в качестве отклика часто используется выход продукта, а позже, когда возможности повышения выхода продукта исчерпаны, используют такие отклики, как энергоемкость, себестоимость продукта и т.д.

Пятое: отклик объекта должен быть *универсальным*, т.е. способным всесторонне характеризовать объект. В частности, технологические параметры недостаточно универсальны: они не учитывают экономику.

Шестое: желательно, чтобы отклик объекта имел *физический смысл*, т.е. следует использовать реальную характеристику объекта, либо стремиться найти идеальную характеристику, сравнимую с реальной.

Седьмое: желательно, чтобы отклик *был простым и легко вычисляемым*.

4.2 Способы формирования обобщенного отклика

Реальные объекты и процессы, как правило, очень сложны. Они часто требуют одновременного учета нескольких, иногда очень многих, параметров. При постановке эксперимента в этих случаях исследователь должен выбрать одну из двух альтернатив:

1) выбрать единственный отклик, который достаточно полно характеризует поведение объекта в рамках конкретной цели исследования, а все прочие характеристики использовать в качестве ограничений;

2) построение обобщенного отклика объекта как некоторой функции от множества первоначально выбранных.

В обоих случаях полезно предварительно исследовать возможность уменьшения числа откликов с применением значения коэффициента корреляции или корреляционного отношения (см. п. 6.2): если при проверке конкретной пары откликов значение коэффициента корреляции или корреляционного отношения окажется близким к единице, то при изучении процесса достаточно рассмотреть только один из них, т.е. исключить из рассмотрения тот, который труднее измерить, или тот, физический смысл которого менее ясен.

Для объединения различных откликов, приходится выбирать для каждого из них некоторую безразмерную шкалу, однотипную для всех объединяемых откликов (это делает их сравнимыми). После выбора шкал осуществляется выбор правила комбинирования исходных частных откликов в обобщенный. Рассмотрим наиболее распространенные способы построения обобщенного отклика.

1. *"Свертка"*. Пусть для каждого из частных откликов $y_1, y_2, ..., y_n$ известно оптимальное значение, к которому нужно стремиться – значение y_{i0}, $i = 1,\dots n$. Тогда разность $y_i - y_{i0}$ можно рассматривать как некоторую меру близости к оптимуму. Для исключения размерности каждого отклика и знака разности используют параметры $\left(\dfrac{y_i - y_{i0}}{y_{i0}}\right)^2$, $i = 1,\dots n$. Очевидно, что наилучшим значением каждого из этих параметров будет 0.

Поскольку на практике различные отклики характеризуют разные результаты функционирования объекта, для каждого из них может быть введен некоторый вес a_i, $i = 1,\dots n$ (для этого обычно используют экспертные оценки) и обобщенный отклик вычисляют как

$$Y = \sum_{i=1}^{n} a_i \cdot \left(\frac{y_i - y_{i0}}{y_{i0}}\right)^2,$$

причем $a_i > 0$, $i = 1,\dots n$ и $\sum_{i=1}^{n} a_i = 1$.

2. *"Обобщенная функция предпочтительности"*. В основе построения этой функции лежит идея преобразования значений частных откликов (как количественных, так и качественных) в безразмерную шкалу предпочтительности, например:

Степень предпочтения	Очень хорошо	Хорошо	Удовлетво-рительно	Плохо	Очень плохо
Шкала предпочтительности	1.0÷0.8	0.8÷0.63	0.63÷0.37	0.37÷0.2	0.2÷0

В этой таблице представлены числа, соответствующие некоторым точкам кривой

$$d = e^{-e^{-y}} = d(y).$$

После выбора шкалы предпочтительности необходимо преобразовать частные отклики в частные функции предпочтительности d_i. Например, для такого отклика, как выход химической реакции область хороших результатов (0.8÷0.63) может соответствовать диапазону 70÷50%, а границе между "удовлетворительно" и "плохо" – 40%. Другими словами, для каждого конкретного отклика необходимо скорректировать значения шкалы предпочтительности и соответствующим образом изменить в функции d выражение в квадратных скобках (для приведенного примера $d_i = e^{-e^{-0.925 y_i}}$.

Обобщенная функция предпочтительности имеет вид:

$$D = \sqrt[n]{\prod_{i=1}^{n} d_i}$$

Способ задания этой функции таков, что если хотя бы одна из $d_i = 0$, то и $D = 0$, а $D = 1$ только тогда, когда все $d_i = 1$. Эта функция весьма чувствительна к малым значениям частных предпочтительностей.

Обобщенная функция предпочтительности является абстрактным построением, но она обладает такими важными свойствами, как адекватность, статистическая чувствительность, эффективность, причем эти свойства не ниже, чем у любого частного отклика. Следовательно, эта функция может использоваться в качестве обобщенного отклика.

После выбора или обобщения отклика объекта исследования необходимо рассмотреть все факторы, которые могут влиять на его функционирование. Если какой-либо существенный фактор окажется неучтенным и не будет контролироваться экспериментатором, это значительно увеличит ошибку опытов. При поддержании этого фактора на определенном уровне может быть получено ложное представление об оптимуме, т.к. нет гарантии, что полученный уровень является оптимальным. С другой стороны большое количество факторов, принятых к рассмотрению, увеличивает число опытов.

4.3 Выбор факторов эксперимента

Фактор – это измеряемая переменная величина, оказывающая влияние на объект исследования. Область определения фактора (его возможных, допустимых значений) может быть непрерывной или дискретной. При планировании эксперимента значения факторов принимаются *дискретными*. Границы областей определения факторов в практических задачах характеризуют технические возможности изменения их значений, либо допустимый режим функционирования объекта.

Факторы подразделяются на количественные, которые можно измерять, и качественные – различные вещества, виды технологий, датчики и приборы, исполнители и т.п.

Качественным факторам не соответствует числовая шкала, но при планировании эксперимента к ним применяют условную порядковую шкалу в соответствии с выбранными уровнями, т.е. производится кодирование. Порядок уровней здесь произволен, но после кодирования он фиксируется.

Основные требования к факторам:

1) *управляемость*, т.е. возможность зафиксировать выбранное значение фактора в течение всего опыта (если экспериментальная установка смонтирована на открытой площадке, то температурой воздуха невозможно управлять, ее можно только контролировать, поэтому температуру воздуха нельзя включать в число факторов);

2) *операциональность* – для определения величины фактора необходимо указать последовательность действий (операций), с помощью которых устанавливаются его конкретные значения, например, если фактором является давление в аппарате, то необходимо указать, в какой точке и с помощью какого прибора оно измеряется и как устанавливается;

3) *необходимая точность измерений* значений факторов, степень которой определяется диапазоном их изменения: в процессах, продолжительность которых измеряется десятками часов, минуты можно не учитывать, а в быстрых процессах приходится учитывать доли секунды;

4) *однозначность*, т.е. фактор не должен быть функцией других факторов, например, если известно, что температура в реакторе должна расти линейно, то в качестве фактора вместо линейной функции можно использовать тангенс угла наклона ее графика;

5) *совместимость*, т.е. при одновременном изменении нескольких факторов, все их комбинации должны быть осуществимы и безопасны, аесли обеспечить совместимость не удается, следует разбить области определения факторов на подобласти и решать несколько задач;

6) *независимость*, т.е. возможность установления значения фактора на любом уровне вне зависимости от уровней других факторов. Заметим, что если это условие невыполнимо, планировать эксперимент невозможно.

Уровнями варьирования фактора называются его количественные или качественные состояния, выбранные для эксперимента. В планировании эксперимента значения факторов, соответствующие определенным уровням их варьирования, кодируются. *Интервалом варьирования* фактора называется разность между двумя его значениями, принятыми при кодировании за +1 и –1.

При выборе области определения факторов особое внимание уделяют выбору нулевой точки, или *основного уровня*. Выбор нулевой точки эквивалентен определению исходного состояния объекта исследования, а оптимизация связана с улучшением его состояния по сравнению с состоянием в нулевой точке.

Если в результате анализа и формализации априорной информации удалось определить наилучшее значение отклика объекта, то в качестве основных уровней факторов принимаются их значения, сочетанию которых соответствует наилучшее значение отклика, а если при постановке задачи области определения факторов задаются явно, то за нулевую точку принимается совокупность центров этих областей.

Интервалы варьирования выбирают так, чтобы значения факторов, соответствующие уровням +1 и –1, были отличимы от значения, соответствующего основному уровню, т.е. величина интервала варьирования должна быть больше удвоенной квадратичной ошибки фиксирования данного фактора. Необходимо помнить, что чрезмерное увеличение интервалов варьирования нежелательно, т.к. это может привести к снижению уровня воспроизводимости опытов, а очень малый интервал варьирования может сильно сузить область эксперимента, и привести к потере оптимума отклика.

Объем эксперимента и эффективность оптимизации отклика объекта определяется количеством отобранных факторов (m) и числа уровней факторов (p): необходимое число опытов

$$N = p^m.$$

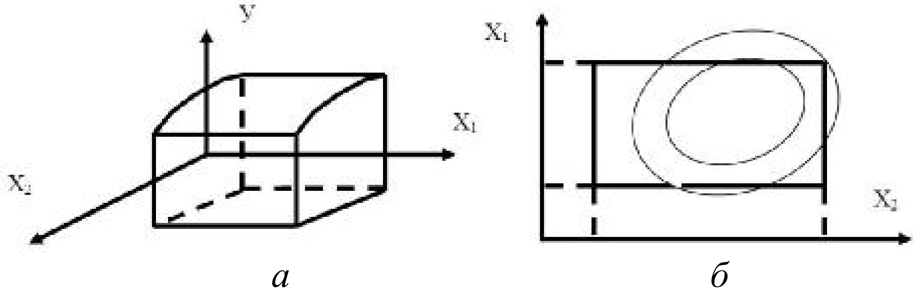
Минимальное число уровней, обычно применяемое на первой стадии работы (при описании объекта исследования линейными уравнениями), равно 2: верхний и нижний уровни, обозначаемые при кодировании через +1 и –1. С увеличением числа уровней повышается чувствительность эксперимента, но одновременно возрастает число опытов: при построении уравнений регрессии второго порядка необходимы 3, 4 или 5 уровней.

Замечание: Необходимо учитывать наличие качественных и дискретных факторов, не применимых при планировании второго порядка, т.к. они не имеют ясного физического смысла для нулевого уровня (применяется преобразование измерительных шкал).

4.4 Выбор вида уравнения регрессии

Выбрать вид уравнения регрессии – значит выбрать один из полиномов, рассмотренных в п. 3.3 или какой-то другой полином, а затем – спланировать и провести эксперимент и определить численные значения коэффициентов этого полинома.

Наглядное, удобно воспринимаемое представление о функции отклика дает ее геометрический аналог – поверхность отклика (ее можно построить, если число факторов не превосходит 2). Пространство, в котором строится поверхность отклика, называется факторным пространством. Оно задается координатными осями, по которым откладываются значения факторов и отклика объекта (рис. 4.1*а*)



а *б*

Рис. 4.1 – Поверхность отклика при двух факторах

а – в трехмерной метрике, *б* – линии равного отклика

Для двух факторов можно не переходить к трехмерному пространству, а ограничиться плоскостью. Для этого достаточно произвести сечения поверхности плоскостями, параллельными плоскости X_1,X_2 (рис. 4.1*б*) и полученные в сечениях линии спроектировать на эту плоскость. Здесь каждая линия соответствует постоянному значению отклика объекта (линия равного отклика).

Требования к уравнению регрессии:

1) *адекватность*: предсказанное с помощью уравнения регрессии значение отклика не отличается от фактического больше, чем на некоторую заранее заданную величину;

2) *простота*: выбор такого полинома регрессии который содержит как можно меньше слагаемых, но удовлетворяет требованию адекватности.

Обычно при планировании эксперимента вначале используют полиномы первой степени, и, если они не удовлетворяют требованию адекватности, поэтапно переходят к полиномам более высоких степеней.

5 СЛУЧАЙНЫЙ ХАРАКТЕР ОТКЛИКА ОБЪЕКТА ИССЛЕДОВАНИЯ

Реальные сложные объекты характеризуются большим количеством ***состояний***, которые определяются ***факторами*** (входными воздействиями на объект), и характеризуются ***откликами*** (выходными величинами). Реальное количество факторов практически не поддается определению, однако в логическую модель объекта исследования экспериментатор вводит ограниченное количество факторов, оказывающих, по его мнению, определяющее влияние на состояние объекта. Чаще всего это вынужденная мера – для упрощения реализации эксперимента и процесса обработки его результатов. Таким образом, с одной стороны, исследователь не включает часть известных ему факторов в эксперимент, а с другой – всегда существуют факторы, о которых он не знает, например, изменение химического состава сырья, ошибки измерения и т.д.

Следовательно, все факторы можно разделить на следующие группы:

1) контролируемые и управляемые факторы (возможна фиксация и изменение их значений);

2) контролируемые, но неуправляемые факторы (возможна только фиксация);

3) неконтролируемые и неуправляемые факторы.

К искажению модели объекта приводит также субъективность процедуры формирования набора факторов, поэтому важно, чтобы логическую модель объекта строил ***совет экспертов***.

Таким образом, модель объекта, разработанная с учетом ограниченного набора факторов, практически всегда является неполной, а реальное поведение объекта складывается под влиянием всех факторов – и включенных в эксперимент, и не включенных, известных экспериментатору, и неизвестных. Тогда значение отклика будет складываться согласно зависимости

$$\boldsymbol{y}=\varphi(x_1,x_2,\ldots,x_m,w_1,w_2,\ldots,w_k),$$

где w_j, j = 1,...,k – неучтенные факторы. Неизвестное нам влияние неучтенных факторов делает отклик объекта $\boldsymbol{y}$ непредсказуемой по значению величиной, а значит – величиной ***случайной***. Таким образом, снятое в эксперименте значение отклика – случайной величины – можно выразить зависимостью

$$\boldsymbol{y} = \varphi(\boldsymbol{x}) + \delta(\boldsymbol{w}),$$

где $\varphi(\boldsymbol{x})$ –функция истинного отклика, отражающая влияние включенных в модель факторов;

$\delta(\boldsymbol{w})$ – функция неучтенных факторов (шума) или просто ***шум***.

5.1 Основные понятия математической статистики

В связи со случайным характером откликов $\boldsymbol{y}$ обработка экспериментальных данных осуществляется с применением аппарата математической статистики – науки, изучающей методы обработки результатов наблюдений массовых случайных явлений, обладающих статистической устойчивостью и закономерностью, с целью выявления этой закономерности.

Функционирование любого объекта реальности моделируется как *случайная величина* – измеримая функция, определённая на вероятностном пространстве. Совокупность всех численных значений случайной величины называется ***генеральной совокупностью***. Участие (или выпадение) какого-то из этих значений в конкретной операционной ситуации непредсказуемо, имеет вероятностный характер и определяется законом распределения вероятностей значений данной случайной величины (соотношение, характеризующее область значений случайной величины и вероятности их появления, см. ниже).

В качестве примеров генеральных совокупностей можно привести среднемесячную температуру июля за сто лет или количество пар мужской обуви, купленной в данном универмаге в обычный будний день за какой-то период времени. Генеральные совокупности принято именовать заглавными латинскими буквами – $\boldsymbol{A}$, $\boldsymbol{X}$, $\boldsymbol{Z}$ и т.д., а конкретные значения величин из данного массива обозначают строчными буквами с индексами : z_1, z_2 и т.д. Генеральные совокупности могут быть конечными или бесконечными, дискретными или непрерывными.

Оперировать с данными всей генеральной совокупности часто невозможно, поэтому их заменяют ***выборками*** (ряд значений данной случайной величины, извлеченных из генеральной совокупности случайным образом). *Представительная выборка*, воспроизводящая закон распределения вероятностей значений случайной величины, обладает такими же свойствами, что и генеральная совокупность, т.е. является как бы ее "мини-моделью". Именно выборки являются объектами работы с данной случайной величиной.

Полученные в опыте значения случайной величины непредсказуемы, но не произвольны: они имеют определенный ***диапазон*** допустимых значений (для непрерывной случайной величины) или ***массив*** допустимых значений (для дискретной случайной величины). Характеристикой случайной величины является ***математическое ожидание*** (оно же – ***генеральное среднее***), которое обозначается как $M\{\boldsymbol{x}\}$, $M\{\boldsymbol{z}\}$ и т.п. Для выборки эквивалентной характеристикой является ***выборочное среднее***, которое является оценкой математического ожидания:

$$\bar{x} = \frac{1}{n} \cdot \sum_{i=1}^{n} x_i ,$$

где n – количество элементов в выборке (объем выборки).

Еще одной характеристикой случайной величины является отклонение ее текущих значений от математического ожидания, т.е. разности $(x_i - M\{\boldsymbol{x}\})$ для генеральной совокупности, либо от выборочного среднего $(x_i - \bar{x})$ для выборок. Общей оценкой этих разностей для всего массива значений является ***дисперсия***, которая для дискретной случайной величины равна

$$\sigma^2 = D\{\boldsymbol{x}\} = \frac{1}{n} \cdot \sum_{i=1}^{n} (x_i - M\{\boldsymbol{x}\})^2 \text{ (для генеральной совокупности)}$$

$$s^2 = \frac{1}{n} \cdot \sum_{i=1}^{n} (x_i - \bar{x})^2 \text{ (для выборки).}$$

<u>Замечание.</u> Величины $M\{\boldsymbol{x}\}$ и σ^2 являются константами как однозначные характеристики всего массива данных, а $\bar{x}$ и s^2 – случайными величинами в связи со случайным характером выборки.

Величины $\bar{x}$ и s^2 выражаются числами и называются точечными оценками $M\{x\}$ и σ^2. Они дополняются интервальными оценками вида: $M\{x\}$ данной случайной величины с фиксированной вероятностью γ лежит в указанном интервале значений случайной величины:

$$p(|x - \bar{x}| < \Delta) = \gamma, \text{ т.е. } \bar{x} - \Delta < x < \bar{x} + \Delta.$$

Этот интервал называется доверительным интервалом, а вероятность γ – доверительной вероятностью. Если выборка содержит значения, не попадающие в доверительный интервал, значит эта выборка непредставительна и должна быть забракована.

Выбор формулы для определения значения Δ зависит от объема выборки и ее вида. Для наиболее распространенной в расчетной практике повторной выборки (случайно отобранный элемент возвращается в генеральную совокупность и может быть отобран повторно):

$$\Delta = t_\gamma \cdot \frac{s}{\sqrt{n}},$$

где значение параметра t_γ зависит от выбранного значения доверительной вероятности, например, при $\gamma = 0.95$ $t_\gamma = 1.96$.

Наиболее полной характеристикой случайной величины является закон распределения ее вероятностей, который связывает конкретное значение случайной величины с вероятностью его появления в опыте. Наиболее распространенным является закон распределения, получивший название ***нормального***. Аналитически этот закон выражается известной формулой Гаусса

$$f(x) = \exp((\bar{x} - M\{x\})^2/2\sigma^2) / \sigma\sqrt{2\pi},$$

где $f(x)$ – плотность вероятностей при данном значении x.

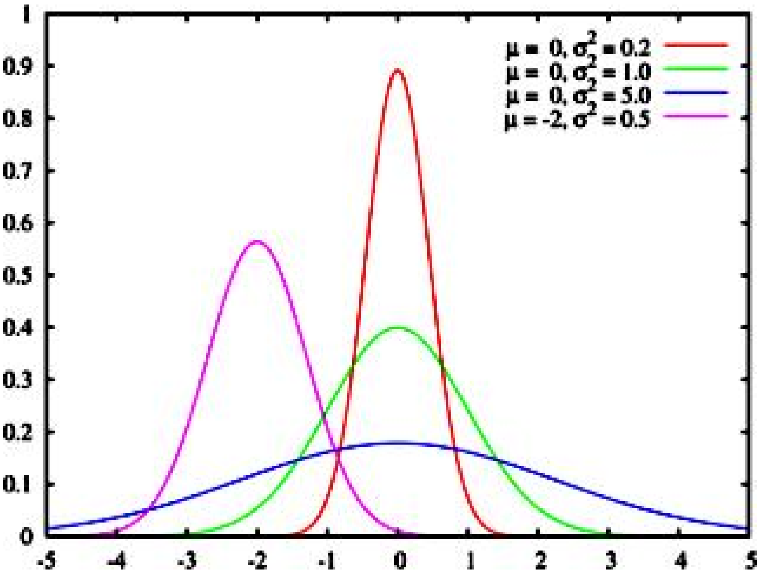


Рисунок 5.1 – Графическое представление нормального распределения случайных величин

Графически эта зависимость представляет собой кривую в форме колокола, которая симметрична относительно центра распределения ($M\{x\}$ – максимум функции $f(x)$), а ее концы уходят в $\pm\infty$, асимптотически приближаясь к оси x, см. рисунок 5.1

Итак, случайная величина есть обособленный поименованный массив чисел, отражающих переменное состояние данного объекта (т.е. являющийся его моделью). Значение отклика в каждой строке таблицы экспериментальных данных есть только одно из значений случайной величины. Каждая из этих случайных величин имеет свои индивидуальные характеристики: математическое ожидание $M\{\boldsymbol{x}\}$, дисперсию σ^2 и закон распределения.

5.2 Ошибки и точность наблюдений в эксперименте

Из всего вышеизложенного следует, что при многократном повторении опыта согласно одной и той же строке таблицы экспериментальных данных (при ***одинаковых*** значениях факторов x) будут получены ***разные*** значения отклика объекта, т.е. за единичным случайным значением отклика объекта y в данной строке таблицы стоит массив случайных величин. Рисунок 5.2 иллюстрирует это положение.

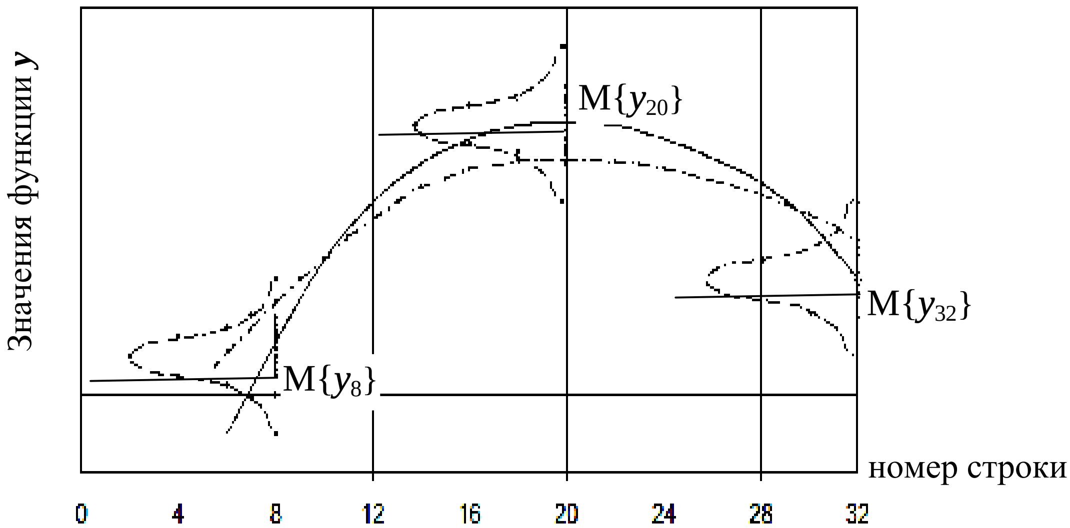


Рисунок 5.2 – Идеальная (сплошная) и экспериментальная (пунктирная) функция ***y***

Дисперсия случайной величины σ^2_y является характеристикой *объекта исследования* и определяется только его природой. Поэтому значение величины σ^2_y одинаково для всех случайных величин во всех строках таблицы данных, а сама дисперсия называется ***дисперсией воспроизводимости*** σ^2_{voz}, т.к. она воспроизводится для всех строк таблицы. Таким образом, графики распределения величин y отличаются только математическими ожиданиями $M\{y\}$, а их дисперсии одинаковы.

Табличное значение величины y является экспериментальной оценкой $M\{y\}$. Надежность оценок зависит от объема выборки и дисперсии случайной величины. На рисунке 5.3 представлены графики законов распределения трех случайных величин при одном значении математического ожидания и различных значениях дисперсии.

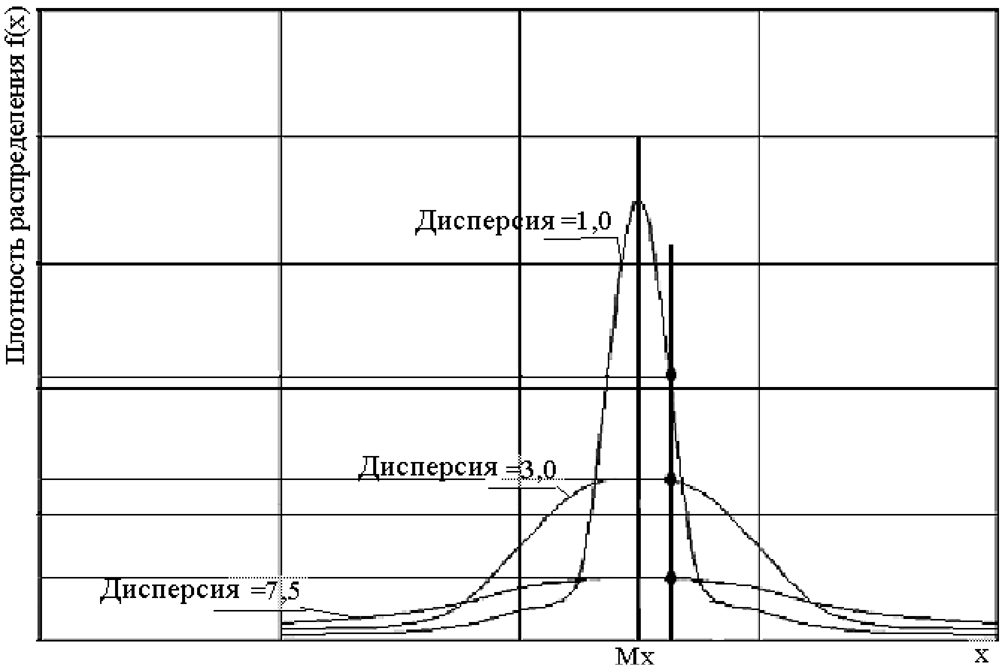


Рисунок 5.3 – Вероятность выпадения данного значения x в зависимости от значения дисперсии

Соотношение $f(x)_{7,5} < f(x)_3 < f(x)_1$ показывает, что чем больше дисперсия, тем более сглажена кривая распределения и тем больше вероятность удаленности экспериментального значения отклика y от значения $M\{\mathbf{y}\}$. Поэтому разность $(y - M\{\mathbf{y}\})$, обусловленную влиянием шума $\delta(\mathbf{w})$, можно рассматривать как "ошибку" экспериментального определения значения отклика y, а дисперсию σ^2_{voz} – как меру этой ошибки.

Таким образом, дисперсия воспроизводимости является мерой ошибки всей процедуры обработки экспериментальных данных. Сравнивая ее по ходу выполнения процедуры с другими показателями меры ошибки, можно оценить степень точности окончательных или промежуточных результатов.

Очевидно, что для получения достоверных результатов с заданной точностью оценки исследуемой величины, нужно провести *не менее* определенного количества наблюдений $\mathbf{n}$. Для его определения используется интервальная оценка математического ожидания этой величины

$$\bar{y} - t_\gamma \cdot \sigma_y / \sqrt{n} < M\{\mathbf{y}\} < \bar{y} + t_\gamma \cdot \sigma_y / \sqrt{n},$$

где $\bar{y}$ – среднее значение случайной величины по выборке.

Выборка имеет определенный размах значений от левой границы $y = q_1$ до правой границы $y = q_2$, т.е. длина интервала значений $\mathbf{y}$ $L = q_2 - q_1$. Чем больше размах значений величины, тем менее достоверны и менее точны выборочные оценки. "Максимум точности" будет достигнут при длине интервала, равной нулю, когда исследуемая величина станет константой. В качестве оценки точности принимают величину $\varepsilon = L/2\sigma_y$, называемую *относительной погрешностью*. Как видно, чем больше интервал значений L, тем меньше точность и больше относительная погрешность.

Если в качестве границ интервала принять $q_1 = \bar{y} - t_\gamma \cdot \sigma_y/\sqrt{n}$, $q_2 = \bar{y} + t_\gamma \cdot \sigma_y/\sqrt{n}$, то $L = 2\cdot(t_\gamma \cdot \sigma_y/\sqrt{n})$, а относительная погрешность $\varepsilon = t_\gamma/\sqrt{n}$, откуда

$$n \geq (t_\gamma / \varepsilon)^2.$$

Для технических объектов "рядового" уровня надежности обычно доверительную вероятность γ принимают равной 0.95, а значение относительной погрешности $\varepsilon = 0.5$. Как указывалось выше, табличное значение $t_{0.95}$ равно 1.96. Тогда минимально необходимое количество наблюдений n будет равно 16, т.е. нужно стремиться провести при значениях факторов данной строки таблицы экспериментальных данных хотя бы 16 наблюдений и вносить в таблицу их среднее значение.

6 ВЗАИМНОЕ ВЛИЯНИЕ СЛУЧАЙНЫХ ВЕЛИЧИН

По таблице экспериментальных данных можно сформировать уравнение регрессии, конкретный вид которого определяет исследователь. При этом всегда возникает вопрос: какова степень соответствия уравнения исходной таблице?

6.1 Стохастическая связь между случайными величинами

Рассмотрим пример: на рисунке 6.1 представлена зависимость между ростом $\boldsymbol{x}$ и весом $\boldsymbol{y}$ студентов-юношей третьего курса одного из вузов России. Посмотрим на поле экспериментальных точек, не обращая пока внимания на график полинома регрессии. Есть ли какая-либо зависимость между величинами $\boldsymbol{x}$ и $\boldsymbol{y}$? Конечно, такая зависимость должна существовать, однако вес человека определяется не только ростом, но и другими факторами, например, окружностью талии. Поэтому, несмотря на очевидную зависимость "вес-рост", мы не можем признать ее однозначной, т.е. функциональной. Видимо, это какая-то другая, нефункциональная зависимость.

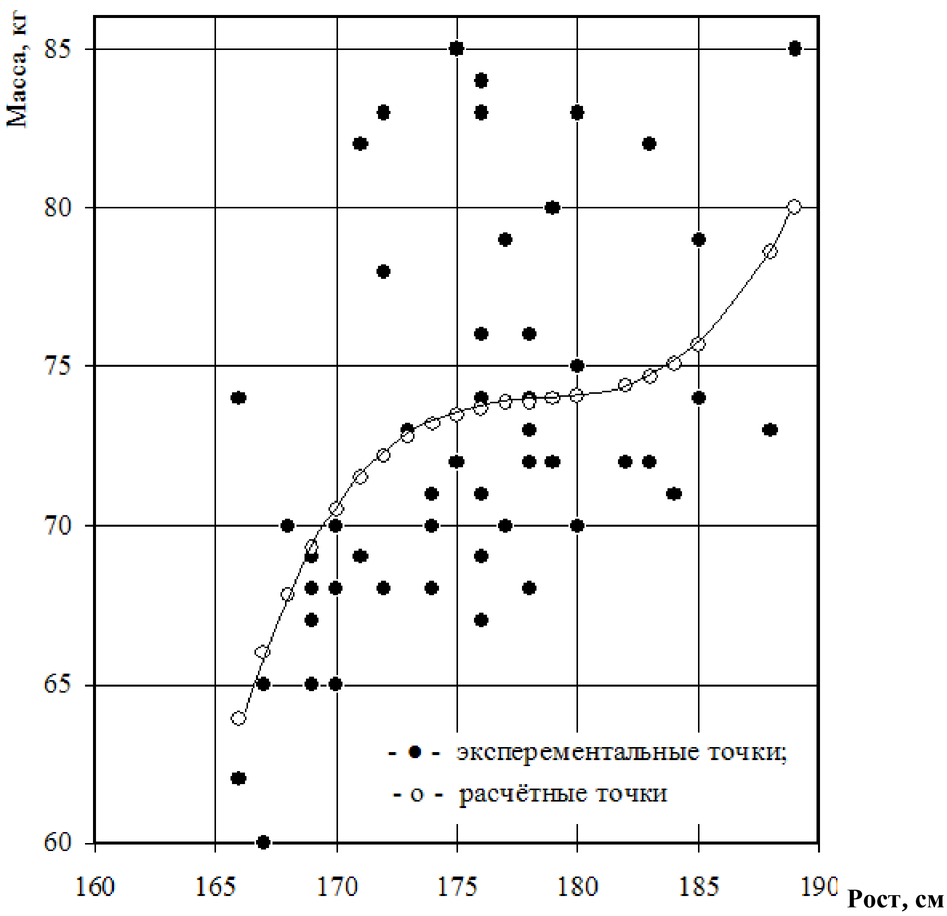


Рисунок 6.1 – Зависимость массы тела студентов от их роста

Полином регрессии, соответствующий экспериментальным данным рисунка 6.1, имеет вид:

$$\boldsymbol{y} = 74.024 + 0.873\boldsymbol{x} - 1.368\boldsymbol{x}^2 + 0.9\boldsymbol{x}^3.$$

Одним из показателей качества таких формул является оценка – насколько близка или далека данная зависимость от реальной функции. Если реальную функцию принять за единицу, то для полученного уравнения этот показатель будет равен 0,513: говорят, что данная зависимость имеет 51.3% функциональности.

Графическое представление подобных зависимостей имеет вид слабо ориентированного облака точек, одному значению аргумента может отвечать несколько значений функции, т.е. появляется ***вероятность*** того или иного значения. Поэтому такой вид связи между величинами называется ***вероятностной*** или ***стохастической*** связью.

В данном конкретном примере такой вид связи обусловлен тем, что в эксперимент и математическую модель объекта в качестве аргументов-факторов включен только рост студентов, хотя очевидно, что существуют и другие факторы, влияющие на функцию, например, размер грудной клетки в сантиметрах. В общем случае стохастическая связь между случайными величинами имеет место тогда, когда они имеют как общие, так и разные аргументы, например $\boldsymbol{y} = f(\bar{u}, \bar{\varepsilon})$ и $\boldsymbol{x} = \varphi(\bar{u}, \bar{\gamma})$. Если влияние общего аргумента ($\bar{u}$) будет нулевым, ***x*** и ***y*** будут независимы, а если влияние разных аргументов ($\bar{\varepsilon}, \bar{\gamma}$)будет нулевым, связь ***x*** и ***y*** будет функциональной. Между этими крайними вариантами лежит бесконечное множество различных состояний по силе стохастической связи. При этом изменение величин ***x*** и ***y*** будет складываться из двух составляющих:

- собственно стохастической под действием общего аргумента $\bar{u}$;
- случайной под действием разных аргументов $\bar{\varepsilon}$ и $\bar{\gamma}$.

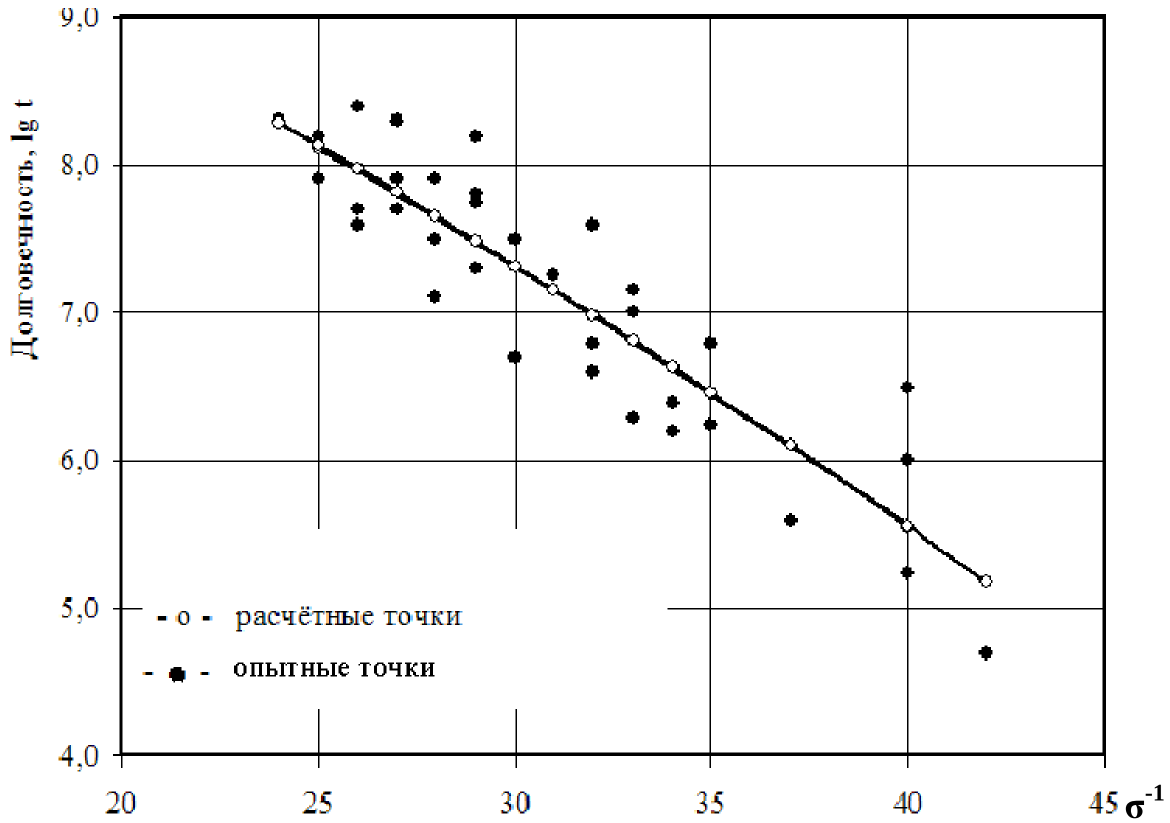


Рисунок 6.2 – Зависимость долговечности образцов жаропрочного сплава от напряжения

Графическое представление сильной стохастической связи – это плотная дорожка точек, т.е. их облако узкое и имеет выраженную направленность. В пределе эта ситуация сводится к линии, т.е. к функции. Слабая связь иллюстрируется рисунком 6.1 – облако размытое, ориентированность направления проявляется слабо. В пределе ситуация сводится к полной хаотичности в расположении точек, т.е. отсутствию зависимости между случайными величинами.

Пример сильной стохастической связи показан на рисунке 6.2. Эта графическая зависимость приближается уравнением

$$\boldsymbol{y} = 1.158 - 0.116\boldsymbol{x} + 0.001\boldsymbol{x}^2.$$

Показатель функциональности этого уравнения равен 0.909 или 90.9%.

Значение случайной величины при фиксированных значениях аргументов не одинаково, его полная характеристика требует учета рассеивания относительно генерального среднего – математического ожидания (например, в виде доверительного интервала), поэтому стохастической называют связь, при которой изменение одной величины вызывает изменение ***закона распределения*** другой.

6.2 Корреляция между случайными величинами

Поскольку дисперсия суммы *независимых* случайных величин равна сумме их дисперсий, т.е. $D\{\boldsymbol{x}+\boldsymbol{y}\}=D\{\boldsymbol{x}\}+D\{\boldsymbol{y}\}$, а $D\{\boldsymbol{z}\}=M\{(\boldsymbol{z}-M\{\boldsymbol{z}\})^2\}$, можно записать

$$D\{\boldsymbol{x}+\boldsymbol{y}\}=M\{[(\boldsymbol{x}+\boldsymbol{y})-M\{(\boldsymbol{x}+\boldsymbol{y})\}]^2\}.$$

Символ математического ожидания суммы разносится по составляющим этой суммы, поэтому

$$\begin{aligned} D\{\boldsymbol{x}+\boldsymbol{y}\} &= M\{(\boldsymbol{x}+\boldsymbol{y}-M\{\boldsymbol{x}\}-M\{\boldsymbol{y}\})^2\}=M\{[(\boldsymbol{x}-M\{\boldsymbol{x}\})+(\boldsymbol{y}-M\{\boldsymbol{y}\})]^2\}= \\ &= M\{(\boldsymbol{x}-M\{\boldsymbol{x}\})^2+2M\{(\boldsymbol{x}-M\{\boldsymbol{x}\})(\boldsymbol{y}-M\{\boldsymbol{y}\})+M\{(\boldsymbol{y}-M\{\boldsymbol{y}\})^2= \\ &= D\{\boldsymbol{x}\}+\boldsymbol{2M\{(x-M\{x\})(y-M\{y\})\}}+D\{\boldsymbol{y}\}. \end{aligned}$$

По сравнению с исходным уравнением появляется дополнительное слагаемое $2M\{(\boldsymbol{x}-M\{\boldsymbol{x}\})(\boldsymbol{y}-M\{\boldsymbol{y}\})\}$которое равно нулю, если случайные величины $\boldsymbol{x}$ и $\boldsymbol{y}$ независимы. При наличии связи между $\boldsymbol{x}$ и $\boldsymbol{y}$, оно принимает какое-то численное значение которое будет тем больше, чем сильнее связь.

Величина $M\{(\boldsymbol{x}-M\{\boldsymbol{x}\})(\boldsymbol{y}-M\{\boldsymbol{y}\})\}$, обозначаемая $\mu_{11}\{\boldsymbol{x}\ \boldsymbol{y}\}$, является показателем силы стохастической связи. На практике используют безразмерную величину – коэффициент корреляции

$$\rho\{\boldsymbol{x}\ \boldsymbol{y}\} = \mu_{11}\{\boldsymbol{x}\ \boldsymbol{y}\}/(\sigma_x \cdot \sigma_y),$$

для выборок – $$r(\boldsymbol{x},\boldsymbol{y}) = \frac{\sum_{i=1}^{n}(x_i-\bar{x})\cdot(y_i-\bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2\cdot\sum_{i=1}^{n}(y_i-\bar{y})^2}},$$

где $\sigma_x \cdot \sigma_y$ – средние квадратичные отклонения,

$\bar{x}, \bar{y}$ – средние арифметические значения x_i, y_i по выборке.

Величины $\rho\{\boldsymbol{x}\ \boldsymbol{y}\}$ и $r(\boldsymbol{x},\boldsymbol{y})$ могут изменяться в пределах [–1,1] и характеризуют не только наличие, но и силу стохастической связи между $\boldsymbol{x}$ и $\boldsymbol{y}$: чем больше абсолютная величина коэффициента корреляции, тем сильнее корреляция между

$\boldsymbol{x}$ и $\boldsymbol{y}$. Максимальной корреляции при $\rho\{\boldsymbol{x}\ \boldsymbol{y}\} = \pm 1$ ("стопроцентная корреляция") будет отвечать наличие функциональной связи между ними.

Для независимых случайных величин коэффициент корреляции равен нулю, но он может быть равен нулю и для некоторых зависимых величин, которые при этом называются *некоррелированными*. Для случайных величин, имеющих нормальное распределение, отсутствие корреляции означает и отсутствие всякой зависимости.

Для определения степени соответствия регрессионного уравнения отклику объекта исследования вычисляют коэффициент корреляции между ними. При $|r(\boldsymbol{y},\boldsymbol{yr})| > 0{,}8$ можно считать, что степень соответствия достаточна.

Замечания. 1. Процедуре получения уравнения регрессии обычно предшествует вычисление коэффициентов корреляции между факторами. При $|r(\boldsymbol{x}_k,\boldsymbol{x}_m)| > 0.95$ любой из двух анализируемых факторов ($\boldsymbol{x}_k$ или $\boldsymbol{x}_m$) можно исключить из рассмотрения как не содержащий дополнительной информации об объекте. Исключают обычно тот фактор, который труднее измерить, или тот, физический смысл которого менее ясен.

2. Коэффициент корреляции характеризует не всякую зависимость, а только линейную, т.е. вероятность того, что при возрастании одной случайной величины другая имеет тенденцию возрастать (или убывать) по линейному закону.

Универсальной характеристикой стохастической связи между случайными величинами, охватывающей все ее виды, является ***корреляционное отношение***

$$\theta = \sqrt{1 - \frac{\sum_{i=1}^{n}(y_i - yr_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}},$$

получаемое из двух дисперсий:

1) остаточной дисперсии уравнения регрессии (характеристика рассеивания наблюдений относительно оценки математической модели объекта исследования)

$$S^2_{\text{ост}} = \frac{1}{n-k+1} SUM_{ost} = \frac{1}{n-k+1}\sum_{i=1}^{n}(y_i - yr_i)^2,$$

где $\boldsymbol{k} + 1$ – количество коэффициентов $\boldsymbol{b}$ в уравнении регрессии;

2) дисперсии вектора $\boldsymbol{y}$ (выборочной дисперсии)

$$S^2_y = \frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})^2,$$

т.е. величина $\theta = \sqrt{1 - \frac{S^2_{\text{ост}} \cdot (n-k+1)}{S^2_y \cdot (n-1)}}$.

Остаточная дисперсия является случайной величиной, так как это функция случайных величин $\boldsymbol{y}$ и $\boldsymbol{yr}$, т.е. $S^2_{\text{ост}}$ имеет математическое ожидание и дисперсию. Можно показать, что $M\{S^2_{\text{ост}}\} = \sigma^2_{voz}$, т.е. $S^2_{\text{ост}}$ – оценка дисперсии воспроизводимости случайной величины $\boldsymbol{y}$.

Остаточная дисперсия так же, как и дисперсия воспроизводимости, является мерой ошибки всей предшествующей процедуры обработки данных и имеет два источника: во-первых, как и σ^2_{voz}, содержит ошибку экспериментального определения значений $\boldsymbol{y}$, а во-вторых – ошибку расчетного определения значений $\boldsymbol{yr}$, т.е. ошибку уравнения регрессии. Таким образом, соотношение σ^2_{voz} и $S^2_{\text{ост}}$ может иметь два результата: если полином регрессии имеет ошибку, остаточная дисперсия будет больше дисперсии воспроизводимости, причем чем больше ошибка полинома, тем больше разница между $S^2_{\text{ост}}$ и σ^2_{voz}; если же полином регрессии адекватен функции истинного отклика объекта исследования $\varphi(\boldsymbol{x})$, то ошибка уравнения отсутствует и $S^2_{\text{ост}} = \sigma^2_{voz}$. Как видно, *сопоставление этих дисперсий позволяет оценить точность полученного уравнения регрессии*.

Значение дисперсии вектора $\boldsymbol{y}$, как и его компонент, определяется двумя факторами: функциональной зависимостью $y = \varphi(x_1, x_2, \ldots, x_m)$ и влиянием функции шума $\delta(\boldsymbol{x})$. Конкретный вид аналитической зависимости $y = \varphi(x_1, x_2, \ldots, x_m)$ неизвестен, но ее табличный вид представляет *объективно* существующую функцию. В значении дисперсии S^2_y эта функция представлена составляющей $\boldsymbol{y}$. Аналогично *субъективная* функция $\boldsymbol{yr}$, которой мы хотим отобразить *объективную* функцию $y = \varphi(x_1, x_2, \ldots, x_m)$, представлена в значении дисперсии $S^2_{\text{ост}}$. Следовательно, сопоставление дисперсий $S^2_{\text{ост}}$ и S^2_y может показать, насколько принятый экспериментатором вид полинома регрессии согласуется с "объективной реальностью" в виде функции истинного отклика $\varphi(\boldsymbol{x})$.

Если уравнение регрессии адекватно функции истинного отклика, т.е. зависимость $y = \varphi(x_1, x_2, \ldots, x_m)$ имеет не стохастический, а функциональный характер, то $\boldsymbol{y} = \boldsymbol{yr}$ и тогда значение корреляционного отношения θ равно единице. Если же связи между величинами $\boldsymbol{y}$ и $\boldsymbol{x}$ нет и зависимость $y = \varphi(x_1, x_2, \ldots, x_m)$ вообще отсутствует, то и в числителе, и в знаменателе выражения для вычисления γ останется только одинаковая составляющая шума $\delta(\boldsymbol{w})$ и значение θ будет равно нулю. Все остальные значения величины θ, промежуточные между "0" и "1", означают переменную "степень функциональности" зависимости между $\boldsymbol{y}$ и $\boldsymbol{x}$. Графически эту "степень функциональности" можно интерпретировать как тесноту размещения точек на графике стохастической зависимости – чем гуще дорожка точек, тем больше значение θ.

Сравнение корреляционных отношений двух разных уравнений регрессии, найденных для одной и той же таблицы экспериментальных данных , позволяет выявить более точное уравнение, при этом разница между значениями θ_1 и θ_2 должна быть статистически значимой.

7 ПРЕДВАРИТЕЛЬНАЯ ОБРАБОТКА ДАННЫХ ЭКСПЕРИМЕНТА И ПРОВЕРКА АДЕКВАТНОСТИ УРАВНЕНИЯ РЕГРЕССИИ

Основной целью предварительной обработки экспериментальных данных является отсеивание грубых погрешностей измерения, подсчета или записи цифрового материала.

7.1 Методика предварительной обработки данных эксперимента

Грубые ошибки при фиксировании результатов эксперимента – это аномальные, сильно выделяющиеся значения в ряду однородных данных. Появление таких значений связано либо с субъективной ошибкой самого экспериментатора, либо с резким нарушением режима проводимых испытаний. Такие значения обычно носят единичный характер и проявляются в одном-двух испытаниях из всей серии. Несмотря на малочисленность, эти значения могут внести существенные искажения в итоговые результаты обработки данных. Поэтому такие аномальные значения должны быть удалены из массива экспериментальных данных, но аномальные значения *не всегда ошибочны*: возможно скачкообразное изменение показателей состояния объекта исследования при плавном изменении значений факторов. Например, при монотонном изменении температуры металлического сплава в нем могут быстро образоваться и так же быстро раствориться новые структурные составляющие (фазы), резко изменяющие свойства сплава. Исключение этих данных может в будущем стать, например, причиной разрушения какой-то конструкции.

Наилучшим выходом из такой ситуации является повторение серии испытаний, которая содержит аномальные результаты и сделать вывод, случаен аномальный результат или нет. Но этот выход не всегда возможен, т.к. "аномальность" обычно обнаруживается при итоговой обработке экспериментального материала, когда повторные эксперименты уже невозможны.

Показателем ошибочности данного наблюдения может служить лишь величина его отклонения от других наблюдений. Сомнительными могут быть крайние отклонения от среднего – как в ту, так и в другую сторону. Если ориентироваться на закон нормального распределения, то такие отклонения симметричны и исследуются одинаково. В случае нормального распределения для единичного значения данной случайной величины x при доверительной вероятности $1-p$ оценкой отсутствия аномалий (однородности) будет соблюдение неравенства

$$\left| x - M\{\boldsymbol{x}\} \right| <= U_{1-p} \cdot \sigma,$$

где $M\{\boldsymbol{x}\}$ – математическое ожидание величины $\boldsymbol{x}$,

σ – ее среднеквадратичное отклонение,

U_{1-p} – табличный квантиль стандартного нормального распределения.

Признаком ошибочности данного значения x является нарушение этого неравенства.

Для выборки объемом n элементов соответствующая доверительная вероятность будет равна $(1 - p)^n$, т.е. вероятность однородности всех n измерений уменьшается с ростом n и при $n \to \infty$ эта вероятность стремиться к нулю. Если x

есть крайний элемент выборки, то доверительной оценке соответствует вероятность $(1-p)^n \cong 1-n \cdot p$. Тогда доверительной вероятности $1-p$ для одного крайнего элемента соответствует оценка

$$\left| x - M\{x\} \right| <= U_{1-p/n} \cdot \sigma.$$

Все вышеизложенное справедливо для случая, когда известны значения $M\{x\}$ и σ. Если они не известны, приходится использовать их выборочные оценки xsr и $S_{\{x\}}$. Тогда для крайнего элемента рабочей статистикой будет условие

$$t_{\text{раб}} = \frac{|x - xsr|}{S_{\{x\}}},$$

которое называется максимальным относительным отклонением и подчиняется распределению Стьюдента. Крайнее значение отбрасывается как грубо ошибочное при условии

$$\frac{|x - xsr|}{S_{\{x\}}} > t_{1-p},$$

где t_{1-p} – табличный квантиль распределения Стьюдента при данном объеме выборки.

После исключения аномального значения статистические характеристики данной выборки пересчитываются для нового объема и новый крайний элемент может быть подвергнут новой проверке. Поскольку при использовании выборочных оценок возникает их смещение относительно оцениваемой величины, в рабочую статистику должна быть введена поправка

$$t_{\text{раб}} = \frac{|x - xsr|}{S_{\{x\}} \cdot \sqrt{\dfrac{n-1}{n}}}.$$

Границы критической зоны $\tau_{p,n}$ (p – нормированное выборочное отклонение) выражаются через квантили распределения Стьюдента $t_{p,n-2}$ по соотношению

$$\tau_{p,n} = \frac{t_{p,n-2} \cdot \sqrt{n-1}}{\sqrt{(n-2) + \left(t_{p,n-2}\right)^2}}.$$

С учетом этого уравнения для выборок большого объема (при $n > 25$) рекомендуют следующую процедуру отсева аномальных данных:

- выбирают значение x_i с максимальным отклонением от среднего;
- вычисляют значение рабочей статистики $t_{\text{раб}}$;
- по таблице t-распределения Стьюдента находят точки $t_{0.05,n-2}$ и $t_{0.001,n-2}$;
- находят критические границы $\tau_{0.05,n}$ и $\tau_{0.001,n}$.

Эти точки ограничивают три зоны: левую до границы $t_{0.05,n-2}$, среднюю между границами $t_{0.05,n-2}$ и $t_{0.001,n-2}$ и правую от границы $t_{0.001,n-2}$. Если значение рабочей статистики попадает в левую зону, крайнее значение не является аномальным. Если оно в средней зоне, то необходим анализ ситуации и выработка дополнительных аргументов в пользу того или иного решения. Если $t_{\text{раб}}$ в правой зоне, крайнее значение безусловно отбрасывается.

Ранее отмечалось, что в качестве j-го значения отклика объекта исследования следует брать среднее значение серии опытов

$$\bar{y}_j = \frac{1}{k_j} \cdot \sum_{i=1}^{k_j} y_{ij} ,$$

где k_j – число опытов по определению j-го значения отклика,

y_{ij} – значение отклика в i-м опыте.

Дисперсию воспроизводимости при определении j-го значения отклика можно оценить значением дисперсии отклика

$$S_j^2 = \frac{1}{k_j - 1} \cdot \sum_{i=1}^{k_j} \left(y_{ij} - \bar{y}_j \right)^2 .$$

Наличие значений S^2_j дает возможность проверить соблюдение условия применимости процедуры регрессионного анализа: равенства дисперсий всех значений отклика. Проверка выполнения этого условия осуществляется с помощью критерия Кохрена:

$$G_{\text{р}} = \frac{\max\limits_j \left\{ S_j^2 \right\}}{\sum\limits_{j=1}^{n} S_j^2} ,$$

где n – число строк в таблице экспериментальных данных.

Расчетное значение критерия Кохрена $G_{\text{р}}$ сравнивается с табличным $G_{\text{т}}$ для выбранного уровня доверительной вероятности (обычно 0.95) и числа степеней свободы

$$f = \frac{1}{n} \cdot \sum_{j=1}^{n} k_j - 1.$$

Если $G_{\text{р}} \le G_{\text{т}}$, то опыты воспроизводимы (гипотеза о равенстве дисперсий воспроизводимости не отвергается). В противном случае следует попытаться добиться выполнения условия воспроизводимости путем выявления и устранения источников нестабильности откликов объекта исследования, например, использовать более точные средства и методы измерений. Иногда бывает достаточно повторить неудавшуюся серию опытов.

7.2 Проверка адекватности уравнения регрессии

Эту проверку можно осуществить лишь в том случае, когда число опытов n превосходит число определяемых коэффициентов уравнения регрессии (k+1) хотя бы на 1.

В качестве оценки дисперсии воспроизводимости определения всех значений отклика объекта исследования, входящих в таблицу экспериментальных данных, используется среднее арифметическое дисперсий всех значений отклика:

$$S_{\{y\}}^2 = \frac{1}{n} \cdot \sum_{j=1}^{n} S_j^2 ,$$

а число степеней свободы этой дисперсии равно сумме чисел степеней свободы дисперсий всех откликов

$$f_{\{y\}} = \sum_{j=1}^{n} \left(k_j - 1\right).$$

Поскольку коэффициенты полинома регрессии вычислены по результатам эксперимента, а эти результаты являются случайными величинами, то случайными величинами являются и сами коэффициенты ***b***. Предполагается, что оценки дисперсий всех коэффициентов равны друг другу и определяются по формуле:

$$S^2_{\{b\}} = \frac{S^2_{\{y\}}}{f_{\{y\}} + n}.$$

Для выявления значимости коэффициентов полинома регрессии для каждого из них рассчитывается значение критерия Стьюдента:

$$t_{\text{р}} = \frac{|b|}{\sqrt{S^2_{\{b\}}}},$$

которое сравнивается с табличным значением $t_{\text{таб}}$ для выбранного уровня доверительной вероятности (обычно 0.95) и числа степеней свободы $f_{\{y\}}$. Если $t_{\text{р}} < t_{\text{таб}}$, то соответствующий коэффициент полинома регрессии может быть приравнен нулю, т.е. соответствующее слагаемое может быть исключено из полинома, а *значения оставшихся коэффициентов необходимо определить заново*.

После проверки значимости коэффициентов и получения окончательного вида полинома регрессии проверяется степень его пригодности – ***адекватности*** объекту исследования. Для этого используются остаточная дисперсия S^2_{ost} уравнения регрессии и дисперсия воспроизводимости σ^2_{vos} экспериментальных данных.

Поскольку обе эти дисперсии являются случайными величинами, сравнивать их нужно с учетом рассеяния и с использованием интервальных оценок, что позволяет установить – ***значимо ли статистически*** различие между ними. Эта значимость проверяется по критерию Фишера: ошибка уравнения регрессии признается значимой, если

$$\frac{S^2_{ost}}{\sigma^2_{vos}} > F_{1-p},$$

где F_{1-p} – значение табличного квантиля распределения Фишера при принятой доверительной вероятности p и степенях свободы $m1 = n - k + 1$, $m2 = +\infty$;

$k + 1$ – количество коэффициентов в полиноме регрессии.

Для учебных расчетов при $p = 0.95$ и $n = 50$ критической границей доверительного интервала ориентировочно можно считать $F_{1-p} = 1.5$: если отношение дисперсий не превосходит 1.5, то они статистически неразличимы, т.е. их можно считать находящимися в одном доверительном интервале, а полином регрессии – адекватным функции истинного отклика $\varphi(x)$. Факт статистической незначимости различия между S^2_{ost} и σ^2_{vos} является ***абсолютным*** показателем точности найденного уравнения регрессии, т.е. того факта, что найденное уравнение следует "принять в эксплуатацию". Если же отношение дисперсий превосходит 1.5, уравнение регрессии имеет ошибку и необходимо взвесить – приемлем ли уровень этой ошибки или нужно искать другое уравнение.

Такую оценку точности уравнения регрессии можно осуществить только при известном значении дисперсии воспроизводимости. Если σ^2_{vos} неизвестна, можно использовать ее оценку – значение $S^2_{\{y\}}$, либо приходится прибегать к сравнительным критериям качества для нескольких альтернативных полиномов с выбором наиболее точного. Статистическую значимость различия дисперсий альтернативных полиномов проводят по условию

$$\frac{S^2_{ost\_i}}{S^2_{ost\_j}} > F_{1-p},$$

где i,j – номера сравниваемых полиномов, причем в числитель помещается большая по значению дисперсия.

В вычислительной практике оценивают отклонение расчетных значений откликов yr_j от результатов эксперимента $\bar{y}_j$ по критерию Фишера следующим образом:

1) в полученное уравнение регрессии построчно подставляются все значения факторов, соответствующие условиям каждого опыта, и определяются значения yr_j, $j = 1,...,n$;

2) вычисляется остаточная дисперсия полинома регрессии

$$S^2_{\text{ост}} = \frac{1}{n-k+1} \cdot \sum_{j=1}^{n} \left(\bar{y}_j - yr_j\right)^2,$$

где $k + 1$ – количество коэффициентов $\boldsymbol{b}$ в уравнении регрессии;

3) вычисляется оценка дисперсии воспроизводимости определения всех значений отклика объекта исследования $S^2_{\{y\}}$;

4) вычисляется расчетное значение критерия Фишера

$$F_p = \frac{S^2_{ost}}{S^2_{\{y\}}};$$

5) по значениям чисел степеней свободы ($f_{\text{ост}} = n - k + 1$ и $f_{\{y\}}$) и выбранному уровню доверительной вероятности определяем табличное значение критерия Фишера $F_{\text{таб}}$ и сравниваем его с $F_{\text{р}}$. Если $F_{\text{р}} < F_{\text{таб}}$, можно считать, что полученный полином регрессии адекватен, т.е. соответствует результатам опытов, в противном случае рекомендуется перейти к другому уравнению регрессии.

Если полином регрессии адекватен результатам эксперимента, осуществляется его интерпретация: устанавливается мера влияния различных факторов на значение отклика. Чем больше абсолютная величина коэффициента при одиночном значении фактора, тем больше его влияние на отклик. Если коэффициент положителен, то с ростом значения фактора значение отклика увеличивается, а если отрицателен, то уменьшается. Таким же образом устанавливается степень влияния на отклик пар факторов, их троек и т.д.

Таким образом, процедура оценки статистической корректности результатов эксперимента предусматривает:

- вычисление значений коэффициентов корреляции между каждой парой факторов и исключение из рассмотрения факторов, не содержащих дополнительной информации об объекте исследования;

- предварительная обработка экспериментальных данных;
- проверка условия воспроизводимости опытов путем вычисления значения критерия Кохрена;
- получение уравнения регрессии выбранного вида;
- проверка значимости коэффициентов уравнения регрессии по критерию Стьюдента, исключение из уравнения незначимых коэффициентов и пересчет значений оставшихся;
- определение уровня стохастической связи между результатами эксперимента и уравнением регрессии путем вычисления значения корреляционного отношения ($\theta > 0.5$);
- проверка адекватности уравнения регрессии объекту исследования путем вычисления значения критерия Фишера и сравнения с табличным значением.

8 ПОЛНЫЙ И ДРОБНЫЙ ФАКТОРНЫЙ ЭКСПЕРИМЕНТ

Работу по планированию эксперимента начинают со сбора и анализа априорной информации (литературные источники, опрос специалистов) с целью выбора отклика (откликов) объекта и факторов, получения представления о характере поверхности отклика, рекомендаций по условиям проведения испытаний. С учетом априорной информации выбирается экспериментальная область факторного пространства: основные (нулевые) уровни и интервалы варьирования факторов.

При планировании эксперимента значения факторов кодируются путем линейного преобразования координат факторного пространства с переносом начала координат в нулевую точку и приведением интервалов варьирования факторов к отрезкам [–1;+1]:

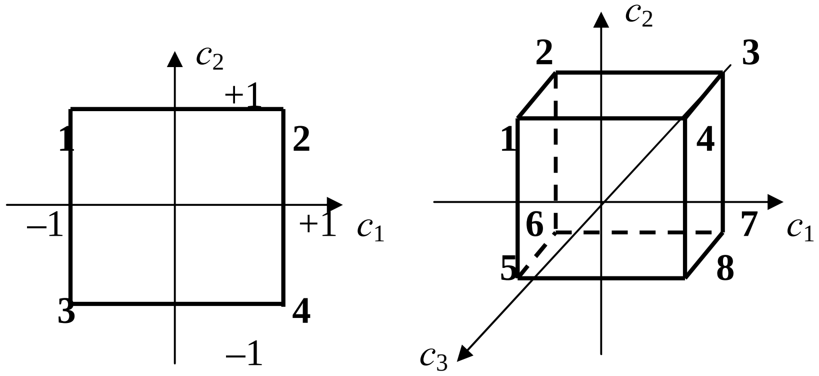
$$c_i = \frac{x_i - x_i^{\min}}{x_i^{\max} - x_i^{\min}},$$

где c_i – кодированное значение i-го фактора (безразмерная величина);

x_i – натуральное значение i-го фактора;

$x_i^{\max}, x_i^{\min}$ – границы интервала варьирования i-го фактора.

Расположение интервалов варьирования кодированных факторов при m = 2 (квадрат) и m = 3 (куб) показано на рис. 8.1. Номерами обозначены граничные точки факторного пространства. По аналогичному принципу располагаются экспериментальные области факторного пространства точки при m > 3.



а) m = 2 б) m = 3

Рисунок 8.1– Области факторного пространства

8.1 Полный факторный эксперимент

Первый этап планирования эксперимента (получение линейного уравнения регрессии) предусматривает два уровня варьирования каждого фактора, т.е. $N = 2^m$. ***Полным факторным экспериментом*** (ПФЭ) называется эксперимент, в котором реализуются все возможные сочетания выбранных уровней факторов. В таблице 8.1 приведены условия ПФЭ типа 2^m при двух факторах, в таблице 8.2 – при трех. Эти таблицы называются матрицами планирования эксперимента, при их заполнении значения уровней факторов обычно обозначают только знаками, а цифру 1 опускают.

Таблица 8.1 – Матрица планирования ПФЭ типа 2^2.

Номер опыта	c_1	c_2	y
1	+	+	y_1
2	+	–	y_2
3	–	+	y_3
4	–	–	y_4

Таблица 8.2 – Матрица планирования ПФЭ типа 2^3.

Номер опыта	c_1	c_2	c_3	y
1	+	+	+	y_1
2	+	–	+	y_2
3	–	+	+	y_3
4	–	–	+	y_4
5	+	+	–	y_5
6	+	–	–	y_6
7	–	+	–	y_7
8	–	–	–	y_8

Каждый столбец в матрице планирования называют вектором-столбцом, а каждую строку – вектором-строкой. Таким образом, в табл. 8.1. два вектора-столбца независимых переменных и один вектор-столбец зависимой переменной, а в таблице 8.2 – соответственно три и один.

При m = 2 все возможные комбинации уровней факторов легко найти перебором, а с ростом числа факторов используют следующий прием, основанный на переходе от матрицы меньшей размерности к матрице большей: исходный план вначале записывается для одного уровня нового фактора, а затем повторяется для другого (см. табл. 8.1, 8.2).

Свойства ПФЭ типа 2^m:

1) симметричность относительно центра эксперимента, т.е. алгебраическая сумма элементов вектора-столбца каждого фактора равна нулю – $\sum_{j=1}^{N} c_{ij} = 0$;

2) "условие нормировки" – сумма квадратов элементов каждого столбца равна числу опытов, т.е. $\sum_{j=1}^{N} c_{ij}^2 = N$ (следствие того, что значения факторов в матрице планирования задаются как +1 и –1);

3) ортогональность матрицы планирования, т.е. сумма почленных произведений любых двух ее векторов-столбцов равна нулю – $\sum_{j=1}^{N} c_{ij} \cdot c_{kj} = 0,\ j \neq k$. Свойство ортогональности позволяет получать независимые оценки коэффициентов уравнения регрессии, т.е. замена нулем любого из них не изменит оценок остальных коэффициентов. Это свойство полезно, когда точный вид уравнения не извес-

тен и требуется по экспериментальным данным отобрать факторы, существенно влияющие на исследуемый отклик;

4) "ротатабельность" матрицы планирования, т.е. ее вектора-строки формируются так, что соответствующие экспериментальные точки находятся на равных расстояниях от центра эксперимента и точность предсказаний значений отклика объекта исследования не зависит от направления.

Выполнение этих условий обеспечивает минимальную дисперсию коэффициентов полинома регрессии, для расчета значений которых выведены следующие соотношения:

- свободный член – $b_0 = \frac{1}{N} \cdot \sum_{j=1}^{N} y_j$;

- коэффициенты при одиночных факторах – $b_i = \frac{1}{N} \cdot \sum_{j=1}^{N} y_j \cdot c_{ij}$;

- коэффициенты при парах факторов – $b_{ik} = \frac{1}{N} \cdot \sum_{j=1}^{N} y_j \cdot c_{ij} \cdot c_{kj}$ и т.д.

8.2 Порядок постановки и оценки точности ПФЭ

Ранее отмечалось, что до начала эксперимента следует рассчитать коэффициенты корреляции между каждой парой отобранных факторов и исключить из рассмотрения один фактор из каждой пары, для которой коэффициент корреляции превосходит 0.8. Однако и после этой операции число факторов может оказаться слишком большим (на практике не рекомендуется использовать более 6 факторов).

Число факторов можно попытаться уменьшить с применением процедуры их ранжирования: расстановки в порядке убывания степени влияния на отклик объекта. Ранжирование факторов осуществляется путем опроса специалистов-экспертов т.е. значением ранга фактора является номер места, отведенного ему экспертом.

Если ломаную фигуру на диаграмме можно соотнести с графиком неравномерного экспоненциального убывания (как на рис. 8.3), то факторы, расположенных ближе к левому флангу ранжированного ряда можно исключить из дальнейшего рассмотрения, отнеся их влияние к шуму (x_3, x_4, x_8). Если же диаграмма равномерна (не содержит явного убывания суммы рангов), то в эксперимент следует включить все факторы.

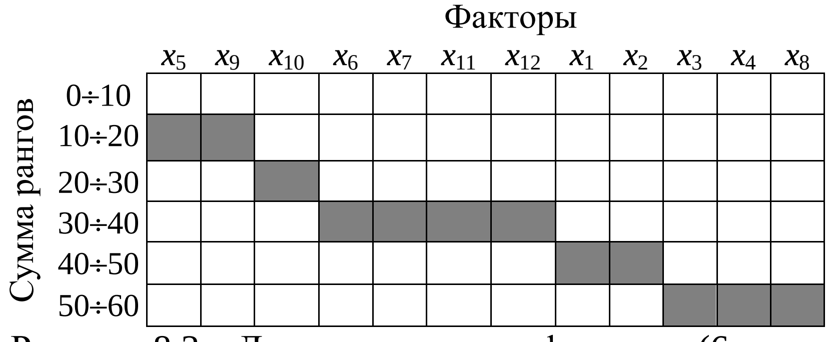


Рисунок 8.3 – Диаграмма рангов факторов (6 экспертов)

Для оценки точности ПФЭ для каждого сочетания уровней факторов (для каждой j-й строки матрицы планирования) проводят K опытов. В результате получают значения $y_{j1}, y_{j2},\ldots, y_{jK}$ исследуемого отклика, которые затем усредняются:

$$\bar{y}_j = \frac{1}{K}\cdot\sum_{k=1}^{K} y_{jk}\ .$$

При этом опыты для определения значений y_{kj} проводят не подряд, а обходят все строки матрицы планирования в первой серии опытов, затем во второй, и так далее до K-й. Для уменьшения влияния внешней среды и неконтролируемых факторов внутри каждой серии строки матрицы планирования обходят случайным образом – *рандомизируют* последовательность опытов с применением генератора случайных чисел. Например, в случае постановки двух серий опытов для ПФЭ типа 2^3 могут быть использованы такие последовательности: 4, 2, 3, 7, 8, 1, 5, 6 (1-я серия) и 2, 4, 6, 8, 5, 7, 3, 1 (2-я серия).

По окончании проведения всех серий измерений всех значений y_{jk} проверяют воспроизводимость опытов с применением критерия Кохрена: для каждой строки матрицы планирования определяется оценка дисперсии отклика

$$S_j^2 = \frac{1}{K-1}\cdot\sum_{k=1}^{K}\left(y_{jk}-\bar{y}_j\right)^2,$$

а затем определяется расчетное значение критерия Кохрена

$$G_{\text{р}} = \frac{\max\limits_{j}\left\{S_j^2\right\}}{\sum\limits_{j=1}^{N} S_j^2}$$

и сравнивают с критическим $G_{\text{кр}}$, которое находят из таблицы распределения Кохрена по числу степеней свободы числителя (K–1), знаменателя (N) и уровню значимости q (обычно 0.05). Если $G_{\text{р}} < G_{\text{кр}}$, гипотеза об однородности дисперсий откликов (воспроизводимости опытов) принимается, в противном случае – отвергается, и тогда эксперимент необходимо повторить, изменив условия проведения (набор факторов, интервал их варьирования, виды и точность измерительных приборов и пр.).

Затем определяются значения коэффициентов уравнения регрессии и проверяется гипотеза их статистической значимости (отличия от нуля). При этом учитывают, что благодаря одинаковой удаленности всех экспериментальных точек, соответствующих строкам матрицы планирования, от центра эксперимента, оценки всех коэффициентов уравнения регрессии независимо от их величины вычисляются с одинаковой погрешностью (при выполнении условия воспроизводимости опытов). Вначале определяют оценку дисперсии воспроизводимости опытов

$$S_y^2 = \frac{1}{N}\cdot\sum_{j=1}^{N} S_j^2$$

и оценку среднеквадратичного отклонения каждого коэффициента (b_j)

$$S_b = \sqrt{\frac{S_y^2}{N}}\,,$$

а затем – расчетное значение критерия Стьюдента для каждого коэффициента

$$t_{\text{р}} = \frac{|b_j|}{S_b}.$$

Его критическое значение $t_{\text{кр}}$ находят из таблицы распределения Стьюдента по числу степеней свободы (N) и уровню значимости q (чаще всего 0.05). Если $t_{\text{р}} > t_{\text{кр}}$, гипотеза о значимости коэффициента b_j принимается, в противном случае коэффициент считается незначимым и приравнивается нулю.

Замечания. 1. При определении статистической значимости коэффициентов уравнения регрессии, полученного по результатам полного факторного эксперимента, количество параллельных опытов не учитывается.

2. Необходимо помнить, что незначимость коэффициента может быть обусловлена неверным выбором интервала варьирования фактора. Поэтому рекомендуется расширить интервал варьирования и провести новый эксперимент.

Заключительным этапом обработки данных ПФЭ является проверка адекватности полученного уравнения регрессии с использованием критерия Фишера: вычисляется дисперсия неадекватности (остаточная дисперсия уравнения регрессии), зависящая от разности между значениями yr_j, рассчитанными с помощью уравнения, и экспериментальными значениями отклика $\overline{y}_j$:

$$S_{\text{ост}}^2 = \frac{1}{N - L + 1} \cdot \sum_{j=1}^{N} \left(\overline{y}_j - yr_j\right)^2,$$

где L – число значимых коэффициентов уравнения, а затем – расчетное значение критерия Фишера

$$F_{\text{р}} = \frac{S_{\text{ост}}^2}{S_y^2}.$$

Критическое значение $F_{\text{кр}}$ находят из таблицы распределения Фишера по числу степеней свободы числителя ($K \cdot (N–L)$), знаменателя ($N \cdot (K–1)$) и уровню значимости q (обычно 0.05). Если $F_{\text{р}} > F_{\text{кр}}$, гипотеза об адекватности отклоняется.

Как правило, вначале проверяют адекватность линейного уравнения регрессии. Если предположение об адекватности подтверждается, то это уравнение фиксируется как окончательное, а если отклоняется – добавляют слагаемые взаимодействия факторов и вновь проверяют гипотезу, и так до тех пор, пока существуют степени свободы (число коэффициентов уравнения меньше числа опытов). Если в результате уравнение все же оказалась неадекватным, значит его форма выбрана неудачно и при данном уровне шума и классе точности измерительных приборов.

Для записи уравнения регрессии *в реальных физических величинах* производят обратный переход от стандартизированного масштаба факторов к натуральному: $x_i = x_i^{\min} + c_i \cdot \left(x_i^{\max} - x_i^{\min}\right)$, после чего записывают окончательный вид уравнения.

8.3 Дробный факторный эксперимент

Количество опытов в полном факторном эксперименте значительно превосходит число определяемых линейных коэффициентов уравнения регрессии, которые обычно имеют наибольшую значимость (коэффициенты взаимодействий факторов часто не значимы). Дробный факторный эксперимент (ДФЭ) позволяет сократить число опытов за счет удаления из матрицы планирования той информации, которая не существенна при построении линейных моделей, причем матрица сохраняет все свои свойства.

Основное правило ДФЭ: из m предварительно отобранных факторов выделяют $m - p$ основных, а остальным присваивают значения, соответствующие ***незначимым*** взаимодействиям основных факторов (не менее двух и не более $m - p$). Следовательно, ДФЭ – это эксперимент типа 2^{m-p}. В качестве примера рассмотрим матрицу планирования ДФЭ типа 2^{3-1}:

№ опыта	c_1	c_2	$c_3 = c_1 \cdot c_2$	y
1	+	+	+	y_1
2	–	+	–	y_2
3	+	–	–	y_3
4	–	–	+	y_4

Как видно, при трех факторах число опытов ДФЭ равно четырем, в то время как число опытов ПФЭ типа 2^3 равнялось бы восьми.

Взаимодействие основных факторов, выбранное для замены дополнительного, называется *генератором плана*, т.к. определяет правило чередования уровней варьирования дополнительного фактора в матрице планирования. В приведенном примере возможными генераторами плана являются взаимодействия $c_1 \cdot c_2$ и $-c_1 \cdot c_2$. При замене $c_1 \cdot c_2$ на $-c_1 \cdot c_2$ матрица планирования ДФЭ типа 2^{3-1} примет вид:

№ опыта	c_1	c_2	$c_3 = -c_1 \cdot c_2$	y
5	+	+	–	y_5
6	–	+	+	y_6
7	+	–	+	y_7
8	–	–	–	y_8

Эти матрицы планирования называются *полурепликами* ПФЭ типа 2^3, т.к. представляют собой верхнюю и нижнюю половины полной матрицы планирования ПФЭ. В ДФЭ могут быть использованы 1/2 (2^{m-1}), 1/4 (2^{m-2}), 1/8 (2^{m-3}) и т.д. части полной матрицы планирования ПФЭ, называемые *дробными репликами*, а ДФЭ иногда называют методом *дробных реплик*.

Вернемся к полурепликам ПФЭ типа 2^3 и посмотрим, каковы будут оценки коэффициентов поверхности отклика объекта β_1, β_2, β_3. Здесь уже не будет тех раздельных оценок ($b_1 \to \beta_1$, $b_2 \to \beta_2$, $b_3 \to \beta_3$), которые были в полном факторном эксперименте. Оценки смешиваются, и для определения эффектов смешения используется *определяющий контраст*: произведение всех факторных векторов-столбцов матрицы планирования. Для верхней полуреплики

$$c_1\cdot c_2\cdot c_3 = col(+1\ \ +1\ \ +1\ \ +1)$$

для нижней $col(-1\ \ -1\ \ -1\ \ -1)$.

Вариант смешения для конкретного фактора определяется умножением соответствующего столбца дробной реплики на определяющий контраст. Так для верхней полуреплики имеем:

$$c_1 = c_1^2\cdot c_2\cdot c_3 = c_2\cdot c_3,\ \text{т.к. } c_1^2 = 1;$$
$$c_2 = c_1\cdot c_2^2\cdot c_3 = c_1\cdot c_3,\ \text{т.к. } c_2^2 = 1;$$
$$c_3 = c_1\cdot c_2\cdot c_3^2 = c_1\cdot c_2,\ \text{т.к. } c_3^2 = 1.$$

Следовательно, коэффициенты уравнения регрессии будут оценками:

$$b_1\to\beta_1+\beta_{23},\ b_2\to\beta_2+\beta_{13},\ b_3\to\beta_3+\beta_{12}.$$

Аналогично для нижней полуреплики:

$$c_1 = -\,c_1^2\cdot c_2\cdot c_3 = -\,c_2\cdot c_3,$$
$$c_2 = -\,c_1\cdot c_2^2\cdot c_3 = -\,c_1\cdot c_3,$$
$$c_3 = -\,c_1\cdot c_2\cdot c_3^2 = -\,c_1\cdot c_2,$$

а коэффициенты уравнения регрессии будут оценками:

$$b_1\to\beta_1-\beta_{23},\ b_2\to\beta_2-\beta_{13},\ b_3\to\beta_3-\beta_{12}.$$

В ДФЭ типа 2^{4-1} генераторами плана могут служить взаимодействия $\pm c_1\cdot c_2$, $\pm c_1\cdot c_3$, $\pm c_2\cdot c_3$, $\pm c_1\cdot c_2\cdot c_3$ (в двух полурепликах). Максимальную разрешающую способность имеют генераторы $c_4 = c_1\cdot c_2\cdot c_3$ и $c_4 = -\,c_1\cdot c_2\cdot c_3$ (линейные эффекты смешаны с эффектами взаимодействия наибольшего порядка). Для определяющих контрастов $1 = c_1\cdot c_2\cdot c_3\cdot c_4$ и $-1 = c_1\cdot c_2\cdot c_3\cdot c_4$ получим:

$$c_1= c_2\cdot c_3\cdot c_4,\quad c_1= -\,c_2\cdot c_3\cdot c_4,\quad c_2 = c_1\cdot c_3\cdot c_4,\quad c_2 = -\,c_1\cdot c_3\cdot c_4,$$
$$c_3 = c_1\cdot c_2\cdot c_4,\quad c_3 = -\,c_1\cdot c_2\cdot c_4,\quad c_4 = c_1\cdot c_2\cdot c_3,\quad c_4 = -\,c_1\cdot c_2\cdot c_3,$$
$$c_1\cdot c_2 = c_3\cdot c_4,\ c_1\cdot c_2 = -\,c_3\cdot c_4,\ c_1\cdot c_3 = c_2\cdot c_4,$$
$$c_1\cdot c_3 = -\,c_2\cdot c_4,\ c_1\cdot c_4 = c_2\cdot c_3,\ c_1\cdot c_4 = -\,c_2\cdot c_3.$$

Такой тип смешивания дает возможность оценивать линейные эффекты совместно с тройными эффектами взаимодействий, а двойные взаимодействия – совместно друг с другом. Здесь коэффициенты уравнения регрессии будут оценками:

$$b_1\to\beta_1+\beta_{234},\ b_2\to\beta_2+\beta_{134},\ b_3\to\beta_3+\beta_{124},\ b_4\to\beta_4+\beta_{123},$$
$$b_{12}\to\beta_{12}+\beta_{34},\ b_{13}\to\beta_{13}+\beta_{24},\ b_{14}\to\beta_{14}+\beta_{23}.$$

Если же в качестве генераторов плана выбрать соотношения $c_4= c_1\cdot c_2$ и $c_4 = -\,c_1\cdot c_2$, то определяющими контрастами будут $1 = c_1\cdot c_2\cdot c_4$ и $-1 = c_1\cdot c_2\cdot c_4$. Варианты смешения:

$$c_1= c_2\cdot c_4,\ c_1 = -\,c_2\cdot c_4,\ c_2 = c_1\cdot c_4,\ c_2 = -\,c_1\cdot c_4,\ c_3 = c_1\cdot c_2\cdot c_3\cdot c_4,\ c_3 = -\,c_1\cdot c_2\cdot c_3\cdot c_4,$$
$$c_4 = c_1\cdot c_2,\quad c_4 = -\,c_1\cdot c_2,\quad c_1\cdot c_3 = c_2\cdot c_3\cdot c_4,\quad c_1\cdot c_3 = -\,c_2\cdot c_3\cdot c_4,$$
$$c_2\cdot c_3 = c_1\cdot c_3\cdot c_4,\quad c_2\cdot c_3 = -\,c_1\cdot c_3\cdot c_4,\quad c_3\cdot c_4 = c_1\cdot c_2\cdot c_3,\quad c_3\cdot c_4 = -\,c_1\cdot c_2\cdot c_3.$$

Разрешающая способность этих генераторов плана ниже, чем у предыдущего примера: здесь линейный коэффициент фактора c_3 зависит от взаимодействия всех факторов: $b_3\to\beta_3+\beta_{1234}$. Выбор такой полуреплики разумен, если имеется априорная информация о большей значимости тройных взаимодействий по сравнению с парными или о незначимости парных взаимодействий $c_1\cdot c_2$, $c_1\cdot c_3$, $c_1\cdot c_4$.

Как видно, выбор конкретной дробной реплики для ДФЭ требует знания значительной априорной информации, что не всегда возможно.

Для ДФЭ стандартизация масштабов факторов, порядок постановки опытов, проверка их воспроизводимости, расчет оценок коэффициентов уравнения регрессии и проверка их статистической значимости, проверка адекватности уравнения регрессии отклику объекта исследования и переход к физическим переменным производятся так же, как и для ПФЭ. Необходимо помнить, что число проводимых опытов должно превышать количество коэффициентов получаемого уравнения регрессии, иначе проверка его адекватности невозможна (нет степеней свободы).

9 ЦЕНТРАЛЬНЫЕ ПЛАНЫ ЭКСПЕРИМЕНТА

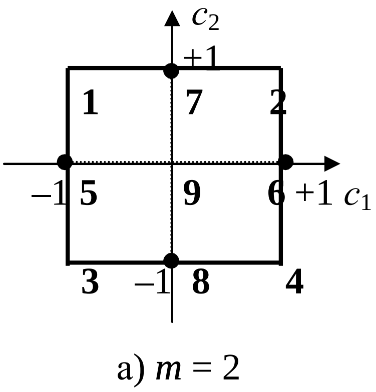
В результате выполнения и обработки результатов ПФЭ и ДФЭ получается уравнение регрессии в виде неполного квадратичного полинома, который содержит линейную часть и сочетания факторов, и, как правило, не имеет выраженного экстремума. Если целью эксперимента является определение значений факторов, соответствующих экстремуму поверхности отклика объекта следует формировать план эксперимента второго порядка. Наибольшее распространение на практике получили ортогональный центральный композиционный план (ОЦКП), ротатабельный центральный композиционный план (РЦКП) и некомпозиционный план Бокса-Бенкена, позволяющие получить уравнение регрессии в виде полного квадратичного полинома.

9.1 Ортогональный центральный композиционный план эксперимента

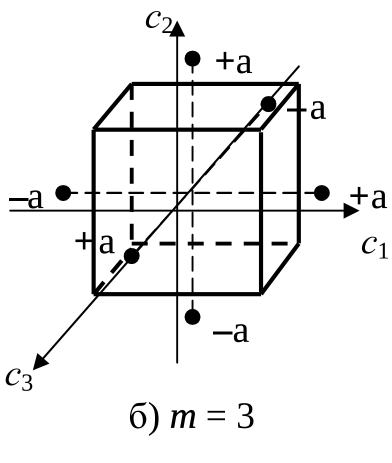
В ОЦКП к опытам ПФЭ или ДФЭ добавляются опыты в так называемых "звездных точках" и в центре плана. Число опытов $N = N_0 + 2\cdot m + 1$, где N_0 – число опытов исходного ПФЭ (2^m) или ДФЭ (2^{m-p}, где p – число генераторов плана). Обычно ОЦКП строится на базе ПФЭ при $m \leq 4$, при $m > 4$ – на базе ДФЭ. Важной особенностью ОЦКП является использование информации, полученной при проведении ПФЭ или ДФЭ.

$2\cdot m$ – это число опытов в "звездных" точках, имеющих координаты: $(\pm\alpha,0,\ldots0)$, $(0,\pm\alpha,0,\ldots0)$,…, $(0,\ldots0,\pm\alpha)$, где α – величина "звездного плеча". Заключительный опыт ставится в центре плана – в точке с координатами $(0,0,\ldots0)$. Иллюстрации ОЦКП для $m = 2$ и $m = 3$ представлены на рис. 9.1.

Выполнение условия ортогональности плана $\sum_{j=1}^{N} c_{ij} = 0$ для векторов столбцов матрицы планирования, соответствующих квадратам факторов требует преобразования их элементов: замены значений c_{ij}^2 на $c_{ij}^2 - a$, причем значение a можно определить из условия симметричности $\sum_{j=1}^{N}\left(c_{ij}^2 - a\right) = \sum_{j=1}^{N} c_{ij}^2 - N\cdot a = 0$, т.е. $a = \frac{1}{N}\cdot\sum_{j=1}^{N} c_{ij}^2$.



а) $m = 2$



б) $m = 3$

Рисунок 9.1– Иллюстрации ОЦКП ("звездные" точки помечены)

В общем случае матрица ОЦКП для трех факторов имеет вид:

	№ опыта	c_1	c_2	c_3	$c_1{\cdot}c_2$	$c_1{\cdot}c_3$	$c_2{\cdot}c_3$	$c_1{\cdot}c_2{\cdot}c_3$	c_1^2-a	c_2^2-a	c_3^2-a	y
Точки ПФЭ 2^3	1	–1	–1	–1	+1	+1	+1	–1	$1-a$	$1-a$	$1-a$	y_1
	2	+1	–1	–1	–1	–1	+1	+1	$1-a$	$1-a$	$1-a$	y_2
	3	–1	+1	–1	–1	+1	–1	+1	$1-a$	$1-a$	$1-a$	y_3
	4	+1	+1	–1	+1	–1	–1	–1	$1-a$	$1-a$	$1-a$	y_4
	5	–1	–1	+1	+1	–1	–1	+1	$1-a$	$1-a$	$1-a$	y_5
	6	+1	–1	+1	–1	+1	–1	–1	$1-a$	$1-a$	$1-a$	y_6
	7	–1	+1	+1	–1	–1	+1	–1	$1-a$	$1-a$	$1-a$	y_7
	8	+1	+1	+1	+1	+1	+1	+1	$1-a$	$1-a$	$1-a$	y_8
"Звездные" точки	9	$-\alpha$	0	0	0	0	0	0	α^2-a	$-a$	$-a$	y_9
	10	$+\alpha$	0	0	0	0	0	0	α^2-a	$-a$	$-a$	y_{10}
	11	0	$-\alpha$	0	0	0	0	0	$-a$	α^2-a	$-a$	y_{11}
	12	0	$+\alpha$	0	0	0	0	0	$-a$	α^2-a	$-a$	y_{12}
	13	0	0	$-\alpha$	0	0	0	0	$-a$	$-a$	α^2-a	y_{13}
	14	0	0	$+\alpha$	0	0	0	0	$-a$	$-a$	α^2-a	y_{14}
Нулевая точка	15	0	0	0	0	0	0	0	$-a$	$-a$	$-a$	y_{15}

Как видно из матрицы, в ОЦКП каждый фактор фиксируется, в общем случае, на пяти уровнях: ($-\alpha$,–1,0,+1,$+\alpha$).

Из условия ортогональности векторов-столбцов (c_1^2-a) и (c_2^2-a) следует:

$$\sum_{j=1}^{N}\left(c_{1j}^2-a\right)\cdot\left(c_{2j}^2-a\right)=N_0\cdot(1-a)^2-4\cdot a\cdot\left(\alpha^2-a\right)+a^2\cdot(N-N_0-4)=0,$$ откуда

$$\frac{1}{N}\cdot\left[N_0-2\cdot a\cdot\left(N_0+2\cdot\alpha^2\right)\right]+a^2=0.$$

Соотношение для определения a из условия симметричности столбца (c_1^2-a):

$$a=\frac{1}{N}\cdot\sum_{j=1}^{N}c_{1j}^2=\frac{1}{N}\cdot\left(N_0+2\cdot\alpha^2\right).$$

Из этих двух уравнений следует:

$$a=\sqrt{\frac{N_0}{N}},\ \alpha=\sqrt{0.5\cdot\left(\sqrt{N\cdot N_0}-N_0\right)}.$$

Например, для ОЦКП при $m=3$:

$$N_0=2^3=8,\ N=8+6+1=15,\ a=\sqrt{8/15}\approx 0.73,\ \alpha=\sqrt{0.5\cdot\left(\sqrt{15\cdot 8}-8\right)}\approx 1.215.$$

По результатам опытов ОЦКП формируется полином:

$$y=b_0+b_1c_1+b_2c_2+b_3c_3+b_{12}c_1c_2+b_{13}c_1c_3+b_{23}c_2c_3+b_{123}c_1c_2c_3+$$
$$+\,b_{11}(c_1^2-a)+b_{22}(c_2^2a)+b_{33}(c_3^2-a),$$

коэффициенты которого определяются по формулам:

$$b_0 = \frac{1}{N} \cdot \sum_{j=1}^{N} y_j \,,\ b_l = \sum_{j=1}^{N} e_{lj} \cdot y_j \Bigg/ \sum_{j=1}^{N} e_{lj}^2 \,,$$

где l – индекс вектора-столбца матрицы ОЦКП,

e_{lj} – значение j-го элемента этого столбца.

Иногда этот полином записывают в виде:

$$y = d_0 + b_1 c_1 + b_2 c_2 + b_3 c_3 + b_{12} c_1 c_2 + b_{13} c_1 c_3 + b_{23} c_2 c_3 + b_{123} c_1 c_2 c_3 + b_{11} c_1^2 + b_{22} c_2^2 + b_{33} c_3^2,$$

где $d_0 = b_0 - a \cdot (b_{11} + b_{22} + b_{33})$.

Чаще всего ОЦКП применяется для построения полиномов регрессии второго порядка. Однако эти планы, кроме двухфакторных, являются избыточными как по числу экспериментальных точек (число экспериментов превосходит число коэффициентов полинома регрессии), так и по числу уровней варьирования факторов (пять вместо трех). Следовательно, ОЦКП при числе факторов более двух может быть использован для построения неполных полиномов регрессии третьего порядка.

Для трех факторов (m = 3) величина звездного плеча α = 1.215, число экспериментальных точек в матрице ОЦКП ($N = 2^m + 2 \cdot m + 1 = 15$) недостаточно для построения полного полинома третьего порядка, но достаточно для построения полинома вида:

$$y = b_0 + b_1 x_1 + b_2 x_2 + b_3 x_3 + b_{12} x_1 x_2 + b_{13} x_1 x_3 + b_{23} x_2 x_3 +$$
$$+ b_{22} x_2^2 + b_{11} x_1^2 + b_{33} x_3^2 + b_{123} x_1 x_2 x_3 + b_{111} x_1^3 + b_{222} x_2^3 + b_{333} x_3^3,$$

т.е. число определяемых коэффициентов полинома равно 14 (на 1 меньше числа опытов)

Для ортогонализации после масштабирования значений факторов вводится замена переменных: $c_i^3 \to c_i^3 - \eta c_i$, i = 1,2,3, т.е. матрица планирования примет вид:

	№ опыта	c_1	c_2	c_3	$c_1 \cdot c_2$	$c_1 \cdot c_3$	$c_2 \cdot c_3$	$c_1 \cdot c_2 \cdot c_3$	$c_1^2 - a$	$c_2^2 - a$	$c_3^2 - a$	$c_1^3 - \eta c_1$	$c_2^3 - \eta c_2$	$c_3^3 - \eta c_3$	y
Точки ПФЭ 2^3	1	–1	–1	–1	+1	+1	+1	–1	$1-a$	$1-a$	$1-a$	$-1+\eta$	$-1+\eta$	$-1+\eta$	y_1
	2	+1	–1	–1	–1	–1	+1	+1	$1-a$	$1-a$	$1-a$	$1-\eta$	$-1+\eta$	$-1+\eta$	y_2
	3	–1	+1	–1	–1	+1	–1	+1	$1-a$	$1-a$	$1-a$	$-1+\eta$	$1-\eta$	$-1+\eta$	y_3
	4	+1	+1	–1	+1	–1	–1	–1	$1-a$	$1-a$	$1-a$	$1-\eta$	$1-\eta$	$1-\eta$	y_4
	5	–1	–1	+1	+1	–1	–1	+1	$1-a$	$1-a$	$1-a$	$-1+\eta$	$-1+\eta$	$1-\eta$	y_5
	6	+1	–1	+1	–1	+1	–1	–1	$1-a$	$1-a$	$1-a$	$1-\eta$	$-1+\eta$	$1-\eta$	y_6
	7	–1	+1	+1	–1	–1	+1	–1	$1-a$	$1-a$	$1-a$	$-1+\eta$	$1-\eta$	$1-\eta$	y_7
	8	+1	+1	+1	+1	+1	+1	+1	$1-a$	$1-a$	$1-a$	$1-\eta$	$1-\eta$	$1-\eta$	y_8
"Звездные" точки	9	$-\alpha$	0	0	0	0	0	0	α^2-a	$-a$	$-a$	$-\alpha^3+\eta\alpha$	0	0	y_9
	10	$+\alpha$	0	0	0	0	0	0	α^2-a	$-a$	$-a$	$\alpha^3-\eta\alpha$	0	0	y_{10}
	11	0	$-\alpha$	0	0	0	0	0	$-a$	α^2-a	$-a$	0	$-\alpha^3+\eta\alpha$	0	y_{11}
	12	0	$+\alpha$	0	0	0	0	0	$-a$	α^2-a	$-a$	0	$\alpha^3-\eta\alpha$	0	y_{12}
	13	0	0	$-\alpha$	0	0	0	0	$-a$	$-a$	α^2-a	0	0	$-\alpha^3+\eta\alpha$	y_{13}
	14	0	0	$+\alpha$	0	0	0	0	$-a$	$-a$	α^2-a	0	0	$\alpha^3-\eta\alpha$	y_{14}
т. 0	15	0	0	0	0	0	0	0	$-a$	$-a$	$-a$	0	0	0	y_{15}

Значение η можно определить из условия ортогональности столбцов матрицы планирования, соответствующих факторам в первой и в третьей степени: $\sum_{j=1}^{N} c_{ij} \cdot \left(c_{ij}^3 - \eta \cdot c_{ij}\right) = \sum_{j=1}^{N} c_{ij}^4 - \left(2^m + 2 \cdot \alpha^2\right) \cdot \eta = 0$, т.е. $\eta = \frac{1}{\left(2^m + 2 \cdot \alpha^2\right)} \cdot \sum_{j=1}^{N} c_{ij}^4 = 1.1284$.

Используя параметры замен переменных, легко получить форму уравнения регрессии:

$$y = b_0 + b_1c_1 + b_2c_2 + b_3c_3 + b_{12}c_1c_2 + b_{13}c_1c_3 + b_{23}c_2c_3 + b_{123}c_1c_2c_3 + b_{11}(c_1^2 - a) + \\ + b_{22}(c_2^2 - a) + b_{33}(c_3^2 - a) + b_{111}(c_1^3 - \eta c_1) + b_{222}(c_2^3 - \eta c_2) + b_{333}(c_3^3 - \eta c_3)$$

или

$$y = d_0 + d_1c_1 + d_2c_2 + d_3c_3 + b_{12}c_1c_2 + b_{13}c_1c_3 + b_{23}c_2c_3 + b_{123}c_1c_2c_3 + \\ + b_{11}c_1^2 + b_{22}c_2^2 + b_{33}c_3^2 + b_{111}c_1^3 + b_{222}c_2^3 + b_{333}c_3^3,$$

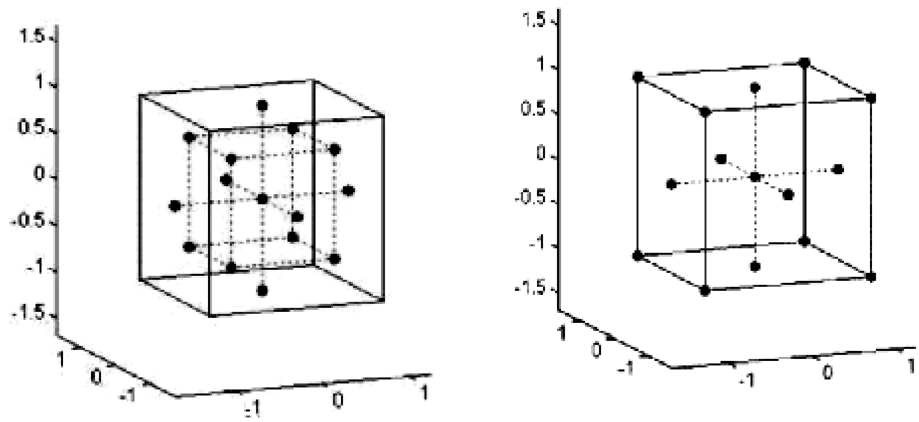
где $d_0 = b_0 - a \cdot (b_{11} + b_{22} + b_{33})$, $d_1 = b_1 - \eta \cdot b_{111}$, $d_2 = b_2 - \eta \cdot b_{222}$, $d_3 = b_3 - \eta \cdot b_{333}$.

Коэффициенты $b_0 \ldots b_{333}$ определяются по тем же соотношениям, что и при формировании полинома регрессии второго порядка. Таков же порядок стандартизации масштабов факторов, постановки опытов и проверки их воспроизводимости, расчета оценок коэффициентов уравнения регрессии и проверки их статистической значимости, проверки адекватности уравнения регрессии отклику объекта исследования и перехода к физическим переменным.

9.2 Модификации ортогонального центрального композиционного плана эксперимента

В практике планирования инженерных экспериментов достаточно часто используются модификации ОЦКП – вписанный ортогональный центральный композиционный план (ВОЦКП) и гранецентрированный центральный композиционный план (ГЦКП).

ВОЦКП представляет собой ОЦКП, масштабированный таким образом, чтобы «звёздные точки» принадлежали граням гиперкуба, т.е. нормированные значения факторов не выходили за пределы отрезков [–1;+1] см. рис. 9.2а. ГЦКП отличается от ОЦКП тем, что звездное плечо $\alpha = 1$, т.е. "звездные точки" лежат в центре граней гиперкуба: на границах отрезков [–1;+1] см. рис. 9.2б.



а *б*

Рисунок 9.2 – Модификации ОЦКП

а – ВОЦКП; *б* – ГЦКП

При использовании ВОЦКП каждый фактор фиксируется на уровнях $(-1,-g,0,+g,+1)$, где $g = 1/\alpha$ для соответствующего количества факторов, а при использовании ГЦКП – на уровнях $(-1,0,+1)$.

Матрица ВОЦКП для трех факторов имеет вид:

	№ опыта	c_1	c_2	c_3	$c_1 \cdot c_2$	$c_1 \cdot c_3$	$c_2 \cdot c_3$	$c_1 \cdot c_2 \cdot c_3$	$c_1^2 - a$	$c_2^2 - a$	$c_3^2 - a$	y
Точки ПФЭ 2^3	1	$-g$	$-g$	$-g$	$+g$	$+g$	$+g$	$-g$	$g^2 - a$	$g^2 - a$	$g^2 - a$	y_1
	2	$+g$	$-g$	$-g$	$-g$	$-g$	$+g$	$+g$	$g^2 - a$	$g^2 - a$	$g^2 - a$	y_2
	3	$-g$	$+g$	$-g$	$-g$	$+g$	$-g$	$+g$	$g^2 - a$	$g^2 - a$	$g^2 - a$	y_3
	4	$+g$	$+g$	$-g$	$+g$	$-g$	$-g$	$-g$	$g^2 - a$	$g^2 - a$	$g^2 - a$	y_4
	5	$-g$	$-g$	$+g$	$+g$	$-g$	$-g$	$+g$	$g^2 - a$	$g^2 - a$	$g^2 - a$	y_5
	6	$+g$	$-g$	$+g$	$-g$	$+g$	$-g$	$-g$	$g^2 - a$	$g^2 - a$	$g^2 - a$	y_6
	7	$-g$	$+g$	$+g$	$-g$	$-g$	$+g$	$-g$	$g^2 - a$	$g^2 - a$	$g^2 - a$	y_7
	8	$+g$	$+g$	$+g$	$+g$	$+g$	$+g$	$+g$	$g^2 - a$	$g^2 - a$	$g^2 - a$	y_8
"Звездные" точки	9	–1	0	0	0	0	0	0	$1 - a$	$-a$	$-a$	y_9
	10	+1	0	0	0	0	0	0	$1 - a$	$-a$	$-a$	y_{10}
	11	0	–1	0	0	0	0	0	$-a$	$1 - a$	$-a$	y_{11}
	12	0	+1	0	0	0	0	0	$-a$	$1 - a$	$-a$	y_{12}
	13	0	0	–1	0	0	0	0	$-a$	$-a$	$1 - a$	y_{13}
	14	0	0	+1	0	0	0	0	$-a$	$-a$	$1 - a$	y_{14}
т. 0	15	0	0	0	0	0	0	0	$-a$	$-a$	$-a$	y_{15}

а матрица ГЦКП – вид:

	№ опыта	c_1	c_2	c_3	y
Точки ПФЭ 2^3	1	–1	–1	–1	y_1
	2	+1	–1	–1	y_2
	3	–1	+1	–1	y_3
	4	+1	+1	–1	y_4
	5	–1	–1	+1	y_5
	6	+1	–1	+1	y_6
	7	–1	+1	+1	y_7
	8	+1	+1	+1	y_8
"Звездные" точки	9	–1	0	0	y_9
	10	+1	0	0	y_{10}
	11	0	–1	0	y_{11}
	12	0	+1	0	y_{12}
	13	0	0	–1	y_{13}
	14	0	0	+1	y_{14}
т. 0	15	0	0	0	y_{15}

Из условия ортогональности векторов-столбцов ($c_1^2 - a$) и ($c_2^2 - a$) матрицы ВОЦКП следует:

$\sum_{j=1}^{N}\left(c_{1j}^2 - a\right)\cdot\left(c_{2j}^2 - a\right) = N_0 \cdot \left(g^2 - a\right)^2 - 4\cdot a\cdot(1-a) + a^2\cdot(N - N_0 - 4) = 0$, откуда после преобразований получается уравнение:

$$N \cdot a^2 - \left(2\cdot N_0 \cdot g^2 + 4\right)\cdot a + N_0 \cdot g^4 = 0.$$

Подставив в это уравнение соотношения для определения значения a из условия симметричности столбца ($c_1^2 - a$)

$$a = \frac{1}{N}\cdot\sum_{j=1}^{N} c_{1j}^2 = \frac{1}{N}\cdot\left(N_0 \cdot g^2 + 2\right),$$

получим уравнение:

$$N_0 \cdot g^4 - \frac{1}{N}\cdot\left(N_0 \cdot g^2 + 2\right)^2 = 0,$$

откуда определяется значение g и далее – значение a. Например, при трех факторах $g \sim 0.823$ (1/1.215), $a = (8\cdot 0.677+2)/15 \sim 0.494$.

Вид полинома регрессии, получаемого при использовании ВОЦКП, а также формулы для определения значений его коэффициентов полностью аналогичны приведенным в п. 9.2 для ОЦКП. Заметим, что применение ВОЦКП недостаточно точно характеризует поведение поверхности отклика на границах области определения факторов, поскольку в точках, соответствующих вершинам гиперкуба, эксперименты не проводятся.

ГЦКП лишен этого недостатка, однако его матрица не обладает свойством ортогональности. Для определения значений коэффициентов полинома регрессии необходимо формировать и решать систему условных уравнений, см. п. 3.4, а исключение из уравнения слагаемых с незначимыми коэффициентами требует повторного формирования и решения системы условных уравнений.

9.3 Ротатабельный центральный композиционный план эксперимента

Ротатабельный центральный композиционный план эксперимента (РЦКП), позволяет получить математическое представление отклика объекта исследования в виде полного полинома второй степени:

$$yr = b_0 + \sum_{i=1}^{m} b_i \cdot x_i + \sum_{i=1}^{m}\sum_{l=i}^{m} b_{il}\cdot x_i \cdot x_l .$$

РЦКП, как и ОЦКП, позволяет сохранить экспериментальную информацию, полученную при реализации ПФЭ или ДФЭ, которую необходимо дополнить опытами в "звездных" точках и в центре плана. Применение РЦКП позволяет получить более точное математическое описание, чем ОЦКП, благодаря увеличению числа опытов в центре плана и способу выбора "звездного плеча" α.

Число опытов при использовании РЦКП

$$N = N_Я + 2\cdot m + N_0,$$

где $N_Я$ – число опытов в ядре плана (ПФЭ или ДФЭ),

m – число факторов,

$2 \cdot m$ – число опытов в в "звездных" точках (с координатами $(\alpha, 0, \dots, 0)$, $(0, \alpha, \dots, 0)$, $(0, 0, \dots \alpha)$ после масштабирования факторов),

N_0 – число опытов в центре плана (в точке, которая после масштабирования факторов будет иметь координаты $(0, 0, \dots, 0)$).

Например, если $m = 3$ и ядром плана является ПФЭ, то $N_Я = 2^3 = 8$, $2 \cdot m = 6$. Число опытов в центре плана выбирается из условия: информация о значениях отклика объекта должна быть неизменной (почти неизменной) для точек сферы единичного радиуса, центр которой находится в центре плана. Для $m = 3$ $N_0 = 6$, т.е. $N = 20$.

Ротатабельность плана (равные расстояния от всех экспериментальных точек до центра плана), обеспечивающая одинаковую точность предсказаний значений отклика объекта исследования для всех точек, достигается выбором значения "звездного плеча" из условий: $\alpha = \begin{cases} 2^{0.25 \cdot m}, \ m < 5 \\ 2^{0.25 \cdot (m-1)}, \ m \geq 5 \end{cases}$, т.е. при $m = 3$ $\alpha = 1.682$.

Матрица планирования РЦКП отличается от матрицы планирования ОЦКП отсутствием тройного взаимодействия факторов и числом опытов в центре плана:

	№ опыта	c_1	c_2	c_3	$c_1 \cdot c_2$	$c_1 \cdot c_3$	$c_2 \cdot c_3$	c_1^2	c_2^2	c_3^2	y
Ядро плана (ПФЭ 2^3)	1	–1	–1	–1	+1	+1	+1	+1	+1	+1	y_1
	2	+1	–1	–1	–1	–1	+1	+1	+1	+1	y_2
	3	–1	+1	–1	–1	+1	–1	+1	+1	+1	y_3
	4	+1	+1	–1	+1	–1	–1	+1	+1	+1	y_4
	5	–1	–1	+1	+1	–1	–1	+1	+1	+1	y_5
	6	+1	–1	+1	–1	+1	–1	+1	+1	+1	y_6
	7	–1	+1	+1	–1	–1	+1	+1	+1	+1	y_7
	8	+1	+1	+1	+1	+1	+1	+1	+1	+1	y_8
"Звездные" точки	9	$-\alpha$	0	0	0	0	0	α^2	0	0	y_9
	10	$+\alpha$	0	0	0	0	0	α^2	0	0	y_{10}
	11	0	$-\alpha$	0	0	0	0	0	α^2	0	y_{11}
	12	0	$+\alpha$	0	0	0	0	0	α^2	0	y_{12}
	13	0	0	$-\alpha$	0	0	0	0	0	α^2	y_{13}
	14	0	0	$+\alpha$	0	0	0	0	0	α^2	y_{14}
Опыты в центре плана	15	0	0	0	0	0	0	0	0	0	y_{15}
	16	0	0	0	0	0	0	0	0	0	y_{16}
	17	0	0	0	0	0	0	0	0	0	y_{17}
	18	0	0	0	0	0	0	0	0	0	y_{18}
	19	0	0	0	0	0	0	0	0	0	y_{19}
	20	0	0	0	0	0	0	0	0	0	y_{20}

Эта матрица ***не обладает*** свойством ортогональности.

При ротатабельном центральном композиционном планировании эксперимента от постановки параллельных опытов можно отказаться и оценивать дисперсию воспроизводимости по экспериментам в центре плана.

Формулы для расчета коэффициентов полинома регрессии и их дисперсий при использовании РЦКП существенно сложнее, чем при использовании ОЦКП (следствие отсутствия свойства ортогональности у матрицы планирования). Сначала определяют значения следующих вспомогательных величин:

$$C = \frac{N}{\sum_{j=1}^{N} c_{ij}^2}\ \forall i,\ B = \frac{m \cdot N}{(m+2)\cdot(N-N_0)},\ A = \frac{1}{2\cdot B\cdot[(m+2)\cdot B - m]},$$

$$S_0 = \sum_{j=1}^{N} y_j,\ S_i = \sum_{j=1}^{N} c_{ij}\cdot y_j,\ S_{il} = \sum_{j=1}^{N} c_{ij}\cdot c_{lj}\cdot y_j,\ Sk_i = \sum_{j=1}^{N} c_{ij}^2\cdot y_j;\ i,l = 1,\ldots,m,\ l \neq i.$$

Коэффициенты рассчитываются по формулам:

$$b_0 = \frac{2\cdot A\cdot B}{N}\cdot\left[S_0\cdot B\cdot(m+2) - C\cdot\sum_{i=1}^{m} Sk_i\right],\ b_i = \frac{C\cdot S_i}{N},\ b_{il} = \frac{C^2\cdot S_{il}}{B\cdot N};\ i,l = 1,\ldots,m,\ l\neq i$$

$$b_{ii} = \frac{A\cdot C}{N}\cdot\left\{Sk_i\cdot C\cdot[B\cdot(m+2) - m] + C\cdot(1-B)\cdot\sum_{i=1}^{m} Sk_i - 2\cdot B\cdot S_0\right\},\ i = 1,\ldots,m.$$

Оценка дисперсии воспроизводимости: $S_y^2 = \frac{1}{N_0 - 1}\cdot\sum_{j=N_Я+2\cdot m+1}^{N}\left(y_j - \bar{y}\right)^2$,

где $\bar{y} = \frac{1}{N_0}\cdot\sum_{j=N_Я+2\cdot m+1}^{N} y_j$ – среднее значение откликов объекта для опытов в центре плана.

Эта оценка найдена для числа степеней свободы ($N_0 - 1$).

Оценки дисперсий коэффициентов полинома регрессии:

$$S_{b_0}^2 = \frac{2\cdot A\cdot B\cdot(m+2)}{N}\cdot S_y^2,\ S_{b_i}^2 = \frac{S_y^2}{N-N_0},\ S_{b_{il}}^2 = \frac{C^2\cdot S_y^2}{N};\ i,l = 1,\ldots,m,\ l\neq i$$

$$S_{b_{ii}}^2 = \frac{A\cdot C^2\cdot S_y^2}{N}\cdot[B\cdot(m+1) - (m-1)],\ i = 1,\ldots,m.$$

Коэффициенты значимы, если выполняются условия:

$$\frac{|b_0|}{\sqrt{S_{b_0}^2}} > t_{кр},\ \frac{|b_i|}{\sqrt{S_{b_i}^2}} > t_{кр},\ \frac{|b_{il}|}{\sqrt{S_{b_{il}}^2}} > t_{кр},\ \frac{|b_{ii}|}{\sqrt{S_{b_{ii}}^2}} > t_{кр};\ i,l = 1,\ldots,m,\ l\neq i.$$

Если параллельные опыты не проводятся, то значение критерия Стьюдента $t_{кр}$ берется при числе степеней свободы ($N_0 - 1$).

Замечание: при использовании РЦКП оценки коэффициентов b_i и b_{il}, $i,l = 1,\ldots,m$, $i \neq l$ не коррелированны с оценками остальных коэффициентов, а оценки b_{ii}, $i = 1,\ldots,m$ – коррелированны с b_0 и между собой, т.е. исключение из полинома регрессии любого из коэффициентов b_{ii} может привести к изменению оценок остальных и b_0.

Проверка адекватности полученного при использовании РЦКП полинома регрессии отклику объекта исследования осуществляется по критерию Фишера:

$$F_{\text{р}} = \frac{S^2_{\text{ост}}}{S^2_y} \le F_{\text{кр}},$$

причем оценка остаточной дисперсии определяется по формуле:

$$S^2_{\text{ост}} = \frac{\sum_{j=1}^{N}\left(y_j - yr_j\right)^2 - S_y^2 \cdot (N_0 - 1)}{N - 0.5 \cdot [(m+2)\cdot(m+1)] - (N_0 - 1)},$$

где yr_j – расчетное значение отклика объекта для j-го опыта.

Число степеней свободы для $S^2_{\text{ост}}$: $N - 0.5 \cdot [(m+2)\cdot(m+1)] - (N_0 - 1)$.

9.4 Некомпозиционные планы

Композиционное планирование воплощает идею «от простого к сложному». Если план первого порядка дает неудовлетворительные результаты, то переходят к плану второго порядка путем проведения экспериментов в «звездных» точках и центре плана. Зачастую плана второго порядка оказывается достаточно для адекватного описания объекта исследования.

Если на основе априорной информации известно, что исследуемый объект описывается полиномом второго порядка, то для получения модели некомпозиционные планы в ряде случаев рациональнее центральных композиционных планов второго порядка.

Например, в случае двух факторов экспериментарациональным является план в виде правильного шестиугольника, см. рис. 9.3. Для обеспечения ротатабельности этого плана в центре проводят 4 эксперимента. Следовательно, общее количество экспериментов составляет 10. Представленный план является более экономичным, чем соответствующий ротатабельный центральный композиционный план второго порядка, который предусматривает проведение 13 экспериментов (4 точки в «ядре» плана, 4 «звездных» точки и 5 экспериментов в центре плана).

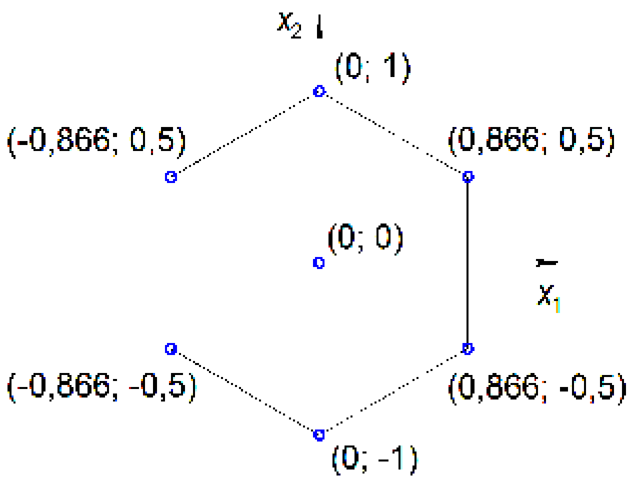


Рисунок 9.3 – Некомпозиционный план для двух факторов эксперимента

Кроме того, по переменной x_1 эксперименты реализуются всего на трех уровнях. РЦКП является 5-уровневым по всем факторам. На практике смена уровня приводит к усложнению эксперимента.

В 1960 г. Дж. Бокс и Д. Бенкен предложили некомпозиционные планы второго порядка, являющиеся определенными выборками строк из полного факторного эксперимента 3^m. На рис. 9.4 показан вид плана для трехфакторного пространства. План предусматривает проведение 15 экспериментов (в центре каждого ребра и 3 в центре плана). Данный план более экономичен, чем соответствующий РЦКП второго порядка, который требует проведения 20 экспериментов.

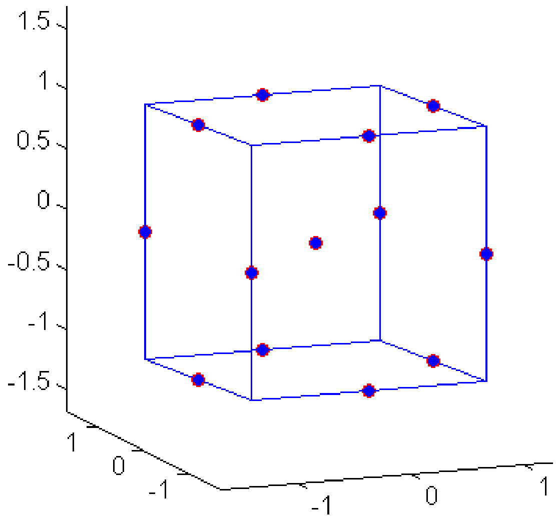


Рисунок 9.4 – Некомпозиционный план Бокса-Бенкена

Сравнение РЦКП (с учетом дробных реплик «ядра») и некомпозиционных планов Бокса-Бенкена свидетельствует о том, что при 3, 4, 6 и 7 факторах последние содержат меньшее количество экспериментов.

Число факторов	3	4	5	6	7
Некомпозиционный план Бокса-Бенкена	15	27	46	54	62
Ротатабельный центральный композиционный план второго порядка	20	31	32	58	92

Ниже представлены матрицы планирования для 3- и 4-факторного планов Бокса-Бенкена. Планы для большего числа факторов приведены в публикации [12], а также могут быть сформированы с помощью функции *boxbehnken* системы Mathcad.

Для расчета коэффициентов полинома второй степени существуют специальные зависимости, однако ввиду из громоздкости в данном пособии они не приводятся. Рекомендуется определять коэффициенты с помощью универсального подхода – метода наименьших квадратов (см. раздел 10.2).

Матрица плана Бокса-Бенкена для 3 факторов

№ опыта	c_1	c_2	c_3
1	1	1	0
2	-1	1	0
3	1	-1	0
4	-1	-1	0
5	1	0	1
6	-1	0	1
7	1	0	-1
8	-1	0	-1
9	0	1	1
10	0	-1	1
11	0	1	-1
12	0	-1	-1
13	0	0	0
14	0	0	0
15	0	0	0

Матрица плана Бокса-Бенкена для 4 факторов

№ опыта	c_1	c_2	c_3	c_4
1	1	1	0	0
2	1	-1	0	0
3	-1	1	0	0
4	-1	-1	0	0
5	0	0	1	1
6	0	0	1	-1
7	0	0	-1	1
8	0	0	-1	-1
9	0	0	0	0
10	1	0	0	1
11	1	0	0	-1
12	-1	0	0	1
13	-1	0	0	-1
14	0	1	1	0
15	0	1	-1	0
16	0	-1	1	0
17	0	-1	-1	0
18	0	0	0	0
19	1	0	1	0
20	1	0	-1	0
21	-1	0	1	0
22	-1	0	-1	0
23	0	1	0	1
24	0	1	0	-1
25	0	-1	0	1
26	0	-1	0	-1
27	0	0	0	0

9.5 Поиск экстремума поверхности отклика объекта

Экстремум поверхности отклика объекта исследования может не совпадать с экстремумом полинома регрессии, найденного в результате проведения и обработки результатов эксперимента. Экстремум может находиться и за пределами предварительно выбранной экспериментальной области факторного пространства: стремление к минимизации интервалов варьирования факторов может привести к потере экстремума. Поэтому после определения экспериментальной области факторного пространства рекомендуется реализовать процедуру поиска экстремума поверхности отклика объекта.

Процедура поиска экстремума (максимума) включает:

1. Определение рабочих и пробных шагов для каждого фактора.

Рабочий шаг можно принять равным половине интервала варьирования фактора: $h_i = 0.5 \cdot \left(x_i^{\max} - x_i^{\min}\right)$, $i = 1,\ldots, m$, а пробный шаг – $\delta_i = K_о \cdot h_i$, где значение коэффициента $K_о < 1$ зависит от возможной точности измерений.

2. Определение направлений изменения значений факторов. Проведение серии пробных опытов: при $x_i = x_i^э$, $i = 1,\ldots, m$ с получением значений y_i, а затем – при $x_i = x_i^э + \delta_i$, $x_j = x_j^э$, $j \neq i$, получение значений yp_i. где $x_i^э$, $i = 1,\ldots, m$ – комбинация значений факторов, соответствующая центру экспериментальной области факторного пространства. Конкретный фактор x_i следует увеличивать, если соответствующее значение конечной разности

$$\Delta_i = \frac{yp\big|_{x_i^э+\delta_i} - y\big|_{x_i^э}}{\delta_i}$$

положительно и уменьшать – если отрицательно.

3. Постановка дополнительных опытов при значениях факторов, увеличенных (уменьшенных) на h_i, $i = 1,\ldots, m$ до тех пор, пока значение отклика объекта увеличивается (при ошибочном выборе значений $x_i^{\max}, x_i^{\min}$ может потребоваться большое число дополнительных опытов, возможно, более эффективным может быть подбор значений h_i: вначале большие, а если значение отклика не увеличивается – более мелкие).

4. Если значение отклика перестало увеличиваться – возврат к п. 2 с лучшей текущей комбинации значений факторов $x_i^э$, $i = 1,\ldots, m$.

5. Завершение процедуры, если первый же дополнительный опыт после выполнения п. 2 не приведет к увеличению значения отклика объекта.

Если в результате выполнения этой процедуры экстремум поверхности отклика объекта оказывается за пределами выбранной экспериментальной области факторного пространства, необходимо скорректировать положение ее границ.

10 ПЛАНОВО-ВЫЧИСЛИТЕЛЬНАЯ ОПТИМИЗАЦИЯ

Для решения задачи оптимизации разработано множество методов, в основе которых лежит направленный перебор возможных вариантов (метод случайного поиска, метод наискорейшего спуска, метод Ньютона и т.д.). Использование классических методов оптимизации предполагает наличие математической модели объекта исследования, которая позволяет ускорить процесс решения задачи и избежать дорогостоящих физических экспериментов. Однако целевую функцию реального объекта редко удается выразить аналитически в виде алгебраических зависимостей от варьируемых параметров. Зачастую поведение объекта описывается дифференциальными уравнениями в частных производных, для решения которых требуются мощные вычислительные ресурсы. В некоторых случаях время моделирования процесса может существенно превышать время его физического протекания. Данное обстоятельство в условиях многомерной оптимизации приводит к огромным затратам машинного времени. В связи с этим целесообразно применение принципиального иного подхода, основанного на теории планирования эксперимента. Идея рассматриваемого метода заключается в замене сложной целевой функции более простой, поиск минимума которой не вызывает затруднений. При этом количество вычислительных экспериментов сводится к минимуму. Например, для построения 16-мерного полинома второй степени согласно некомпозиционному плану Бокса-Бенкена достаточно проведения 396 расчетов, в то время как прямое решение 16-мерной задачи оптимизации не представляется возможным за приемлемое время.

Теория планирования эксперимента изначально разрабатывалась для физических экспериментов. При переходе к экспериментам численным следует учитывать некоторые особенности, главная из которых касается проверки адекватности регрессионной модели.

10.1 Критерий адекватности регрессионной модели

При реализации вычислительных экспериментов дисперсия воспроизводимости равна нулю, т.е. при одинаковых сочетаниях факторов результаты экспериментов не будут различаться. Следовательно, использование рассмотренного ранее критерия Фишера приводит в этом случае к парадоксу: любая модель становится неадекватной. По этой причине рекомендуется в качестве критерия адекватности использовать относительную ошибку модели на области плана:

$$\varepsilon = \frac{\frac{1}{N}\sqrt{\sum_{i=1}^{N}\left(y_i - \mathrm{f}\left(\mathbf{X}^{\mathrm{T}\langle i\rangle}\right)\right)^2}}{y_{\max} - y_{\min}} \cdot 100\%,$$

где y_i – значение численного эксперимента в i-ой точке плана;

$\mathrm{f}\left(\mathbf{X}^{\mathrm{T}\langle i \rangle}\right)$ – значение целевой функции по регрессионной модели в i-ой точке плана;

$y_{\max}$ и $y_{\min}$ – максимальное и минимальное значения целевой функции в точках плана;

$\mathbf{X}^{\mathrm{T}\langle i \rangle}$ – i-й столбец транспонированной матрицы планирования $\mathbf{X}$ в натуральном масштабе (сочетание факторов i-го эксперимента);

N – количество точек плана.

Обычно принимают, что при $\varepsilon \le 5\%$ модель адекватна и ее можно применять для оптимизации целевой функции. В случае $\varepsilon > 5\%$ рекомендуется изменить план эксперимента или уменьшить интервалы варьирования хотя бы некоторых факторов.

Другой распространенный метод проверки адекватности регрессионной модели основан на проведении серии проверочных вычислительных экспериментов в точках гиперкуба области определения, наиболее удаленных от точек плана. Как правило, оценку адекватности проводят по трем проверочным точкам. При этом удобное визуальное представление точности регрессионной модели дает график, по оси абсцисс которого откладываются значения вычислительных экспериментов, по оси ординат – значения уравнения регрессии. Очевидно, что чем ближе точки окажутся к диагонали, тем выше точность регрессионной модели. Решение об ее адекватности принимают на основе расчета средней погрешности в проверочных точках (на рис. 9.5 показаны красным).

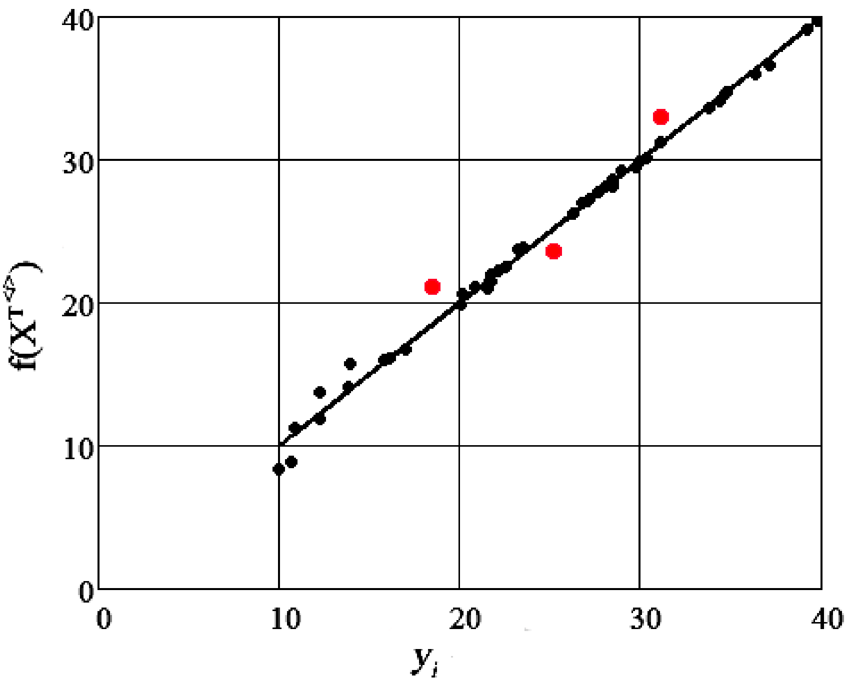


Рисунок 9.5 – Проверка адекватности уравнения регрессии при вычислительном эксперименте

10.2 Определение коэффициентов регрессионной модели по методу наименьших квадратов

Ортогональный план эксперимента и его модификации получили широкое распространение в 1950-х годах благодаря малому объему вычислений, т.к. коэффициенты регрессии определяются независимо друг от друга. С точки зрения распределения информации ортогональное планирование не является оптимальным: информационная функция имеет сложный профиль и зависит от направления движения от центра плана. Данное свойство ортогонального плана является существенным недостатком, поскольку на этапе выбора плана неизвестно, какое направление представляет преимущественный интерес.

Развитие вычислительной техники способствовало распространению на практике неортогонального планирования, стало возможным применение ротатабельных и некомпозиционных планов. Коэффициенты регрессии определяются в этом случае с помощью метода наименьших квадратов, который применим для любого типа планирования. Рассмотрим этот метод на примере плана первого порядка с 4 факторами.

Линейное уравнение регрессии имеет вид

$$\mathrm{f}(x_1, x_2, x_3, x_4) = b_0 + \sum_{i=1}^{4} b_i x_i .$$

Суть метода наименьших квадратов заключается в минимизации суммы квадратов отклонений, т.е.

$$\Phi = \sum_{i=0}^{N-1}\left[y_i - \mathrm{f}\left(\mathbf{X}^{\mathrm{T}<i>}, b_0, b_1, \mathbf{K}, b_5,\right)\right]^2 \to \min .$$

Функция $\Phi(b_0, b_1, ..., b_5)$ имеет единственный минимум, необходимым условием которого является выполнение равенств:

$$\frac{\partial\Phi}{\partial b_0} = 0,\ \frac{\partial\Phi}{\partial b_1} = 0,\ \dots,\ \frac{\partial\Phi}{\partial b_5} = 0.$$

Если слагаемые уравнения регрессии, называемые также базисными функциями, линейно независимы, то после преобразований получим *определенную* систему линейных алгебраических уравнений (СЛАУ) пятого порядка:

$$
\begin{cases}
b_0 N + b_1 \sum\limits_{i=0}^{N-1} x_{1i} + b_2 \sum\limits_{i=0}^{N-1} x_{2i} + b_3 \sum\limits_{i=0}^{N-1} x_{3i} + b_4 \sum\limits_{i=0}^{N-1} x_{4i} = \sum\limits_{i=0}^{N-1} y_i; \\
b_0 \sum\limits_{i=0}^{N-1} x_{1i} + b_1 \sum\limits_{i=0}^{N-1} x_{1i}^2 + b_2 \sum\limits_{i=0}^{N-1} x_{1i} x_{2i} + b_3 \sum\limits_{i=0}^{N-1} x_{1i} x_{3i} + b_4 \sum\limits_{i=0}^{N-1} x_{1i} x_{4i} = \sum\limits_{i=0}^{N-1} y_i x_{1i}; \\
b_0 \sum\limits_{i=0}^{N-1} x_{2i} + b_1 \sum\limits_{i=0}^{N-1} x_{1i} x_{2i} + b_2 \sum\limits_{i=0}^{N-1} x_{2i}^2 + b_3 \sum\limits_{i=0}^{N-1} x_{2i} x_{3i} + b_4 \sum\limits_{i=0}^{N-1} x_{2i} x_{4i} = \sum\limits_{i=0}^{N-1} y_i x_{2i}; \\
b_0 \sum\limits_{i=0}^{N-1} x_{3i} + b_1 \sum\limits_{i=0}^{N-1} x_{1i} x_{3i} + b_2 \sum\limits_{i=0}^{N-1} x_{2i} x_{3i} + b_3 \sum\limits_{i=0}^{N-1} x_{3i}^2 + b_4 \sum\limits_{i=0}^{N-1} x_{3i} x_{4i} = \sum\limits_{i=0}^{N-1} y_i x_{3i}; \\
b_0 \sum\limits_{i=0}^{N-1} x_{4i} + b_1 \sum\limits_{i=0}^{N-1} x_{1i} x_{4i} + b_2 \sum\limits_{i=0}^{N-1} x_{2i} x_{4i} + b_3 \sum\limits_{i=0}^{N-1} x_{3i} x_{4i} + b_4 \sum\limits_{i=0}^{N-1} x_{4i}^2 = \sum\limits_{i=0}^{N-1} y_i x_{4i}.
\end{cases}
$$

Полученную систему можно записать в матричной форме:

$$\mathbf{Mb} = \mathbf{c}.$$

Известно, что симметричная матрица **M** часто имеет очень большое число обусловленности. Для решения плохо обусловленных СЛАУ могут быть использованы метод сингулярного разложения (SVD) или метод *QR*-разложения. Однако, как показывают практические расчеты, при относительно небольших порядках СЛАУ (менее 100) высокую точность решения обеспечивает более простой метод LUP-разложения, являющийся стандартным методом решения СЛАУ в системе Mathcad.

Аналогичным образом могут быть найдены коэффициенты регрессии для плана любого другого порядка и типа. При этом нетрудно обнаружить закономерность в чередовании элементов матрицы **M**, что позволяет автоматизировать процесс ее составления.

Минимум полученного уравнения регрессии может быть найден одним из численных методов поиска экстремума с ограничениями. Поскольку целевая функция представляет собой полином (обычно второй или третьей степени), то ее минимум можно найти аналитически. Если глобальный экстремум окажется за пределами гиперкуба области определения, то необходимо искать минимумы на его гранях и вершинах.

Экстремум исходной целевой функции может быть пропущен на этапе оптимизации из-за погрешности регрессионной модели. По этой причине часто применяется подход, при котором область определения разбивается на более мелкие подобласти (например, 3-мерный куб разбивается на 8 кубов). Далее производится поиск минимума в каждой из этих подобластей. Такой подход обеспечивает получение альтернативных решений, которые могут незначительно различаться по критерию оптимальности. На заключительном этапе проводятся проверочные численные эксперименты при нескольких лучших сочетаниях факторов, и выбирается единственное решение.

11 ЛАБОРАТОРНЫЙ ПРАКТИКУМ

Лабораторный практикум включает задания, методические указания по выполнению и контрольные вопросы к четырем лабораторным работам:

1) Формирование полиномов регрессии заданной конфигурации и оценка степени их корреляции с таблицей данных эксперимента;

2) Предварительная обработка данных эксперимента, формирование и проверка адекватности уравнения регрессии;

3) Полный/дробный факторный эксперимент и обработка его результатов

4) Центральные композиционные планы эксперимента и обработка результатов их применения.

Все работы должны быть выполнены в системе инженерных расчетов Mathcad. Расчетные файлы должны включать необходимые комментарии к исходным данным, расчетным формулам и результатам расчетов.

11.1 Лабораторная работа № 1. Формирование полиномов регрессии заданной конфигурации и оценка степени их корреляции с таблицей данных эксперимента

<u>Задание</u>: сформировать по заданной таблице экспериментальных данных, см. табл. 11.1, два полинома регрессии из нижеприведенного перечня путем минимизации остаточной суммы системы условных уравнений.

Перечень альтернативных уравнений регрессии:

1) неполное квадратичное уравнение, содержащее линейную часть и парные сочетания факторов;

2) полное квадратичное уравнение;

3) неполное кубическое уравнение, содержащее линейную часть, парные и тройные сочетания факторов;

4) полное кубическое уравнение.

Таблица 11.1 Варианты заданий к лабораторной работе № 1

<table>
<tr><th>№ варианта</th><th colspan="9">Таблицы данных эксперимента</th><th>Номера полиномов</th></tr>
<tr><td>1</td><td colspan="9">2</td><td>3</td></tr>
<tr><td rowspan="6">1</td><td colspan="9">Факторы</td><td rowspan="6">1,2</td></tr>
<tr><td>x_1:</td><td>–10</td><td>–8</td><td>–6</td><td>–4</td><td>–2</td><td>0</td><td>2</td><td>4</td></tr>
<tr><td>x_2:</td><td>–8</td><td>–5</td><td>–2</td><td>1</td><td>4</td><td>7</td><td>10</td><td>13</td></tr>
<tr><td>x_3:</td><td>–7</td><td>–5</td><td>–3</td><td>–1</td><td>1</td><td>3</td><td>5</td><td>7</td></tr>
<tr><td colspan="9">Отклик объекта</td></tr>
<tr><td>y:</td><td>6.25</td><td>4.08</td><td>3.85</td><td>3.44</td><td>0.81</td><td>–0.12</td><td>–3.11</td><td>–2.48</td></tr>
<tr><td rowspan="6">2</td><td colspan="9">Факторы</td><td rowspan="6">3,4</td></tr>
<tr><td>x_1:</td><td>–6</td><td>–4</td><td>–2</td><td>0</td><td>2</td><td>4</td><td>6</td><td>8</td></tr>
<tr><td>x_2:</td><td>–6</td><td>–3</td><td>–2</td><td>0</td><td>1</td><td>3</td><td>5</td><td>7</td></tr>
<tr><td>x_3:</td><td>–9</td><td>–6</td><td>–3</td><td>0</td><td>3</td><td>6</td><td>9</td><td>12</td></tr>
<tr><td colspan="9">Отклик объекта</td></tr>
<tr><td>y:</td><td>–3.71</td><td>–2.18</td><td>–0.54</td><td>5.34</td><td>12.87</td><td>25.71</td><td>42.95</td><td>78.63</td></tr>
</table>

Продолжение таблицы 11.1

1	*2*									*3*
3	Факторы									1,3
	x_1:	–11	–7	–4	–1	2	5	8	11	
	x_2:	–11	–8	–5	–2	1	4	7	10	
	x_3:	–8	–6	–4	–2	0	2	4	6	
	Отклик объекта									
	y:	–4.44	–2.86	0.54	1.99	2.43	4.21	5.63	6.26	
4	Факторы									2,4
	x_1:	–10	–7	–4	–1	1	3	5	9	
	x_2:	–9	–5	–2	0	3	6	9	11	
	x_3:	–11	–9	–6	–3	0	4	7	10	
	Отклик объекта									
	y:	–0.25	–0.11	1.06	3.62	8.49	15.32	37.41	80.42	
5	Факторы									1,4
	x_1:	–12	–9	–7	–4	0	2	5	8	
	x_2:	–7	–4	–1	2	4	7	10	12	
	x_3:	–10	–6	–3	0	2	6	9	11	
	Отклик объекта									
	y:	81.23	44.23	16.47	9.4	4.14	1.47	–0.28	–1.35	
6	Факторы									2,3
	x_1:	0	3	7	9	13	16	19	23	
	x_2:	–1	1	3	5	7	9	11	13	
	x_3:	3	6	9	12	15	18	21	24	
	Отклик объекта									
	y:	–9.52	–7.46	–5.48	–3.96	–2.57	–1.03	0.85	1.63	
7	Факторы									1,2
	x_1:	–14	–10	–7	–3	0	4	9	12	
	x_2:	–13	–11	–10	–6	–2	3	7	10	
	x_3:	–16	–14	–11	–7	–2	2	6	9	
	Отклик объекта									
	y:	14.2	8.74	1.67	–3.76	–4.87	–1.14	2.39	7.34	
8	Факторы									3,4
	x_1:	0	4	8	12	16	20	24	28	
	x_2:	0	5	10	15	20	25	30	35	
	x_3:	0	3	6	9	12	15	18	21	
	Отклик объекта									
	y:	8.45	6.49	3.77	–0.96	–2.17	–1.04	0.59	2.34	
9	Факторы									1,3
	x_1:	–10	–8	–6	–4	–2	0	2	4	
	x_2:	–8	–5	–2	1	4	7	10	13	
	x_3:	–7	–5	–3	–1	1	3	5	7	
	Отклик объекта									
	y:	–6.25	–4.08	–3.85	–3.44	–0.81	0.12	3.11	2.48	

Продолжение таблицы 11.1

1	*2*									*3*
10	Факторы									2,4
	x_1:	–6	–4	–2	0	2	4	6	8	
	x_2:	–6	–3	–2	0	1	3	5	7	
	x_3:	–9	–6	–3	0	3	6	9	12	
	Отклик объекта									
	y:	3.71	2.18	0.54	–5.34	–12.87	–25.71	–42.95	–78.63	
11	Факторы									1,4
	x_1:	–11	–7	–4	–1	2	5	8	11	
	x_2:	–11	–8	–5	–2	1	4	7	10	
	x_3:	–8	–6	–4	–2	0	2	4	6	
	Отклик объекта									
	y:	4.44	2.86	–0.54	–1.99	–2.43	–4.21	–5.63	–6.26	
12	Факторы									2,3
	x_1:	–10	–7	–4	–1	1	3	5	9	
	x_2:	–9	–5	–2	0	3	6	9	11	
	x_3:	–11	–9	–6	–3	0	4	7	10	
	Отклик объекта									
	y:	0.25	0.11	–1.06	–3.62	–8.49	–15.32	–37.41	–80.42	
13	Факторы									1,2
	x_1:	–12	–9	–7	–4	0	2	5	8	
	x_2:	–7	–4	–1	2	4	7	10	12	
	x_3:	–10	–6	–3	0	2	6	9	11	
	Отклик объекта									
	y:	–81.23	–44.23	–16.47	–9.4	–4.14	–1.47	0.28	1.35	
14	Факторы									3,4
	x_1:	0	3	7	9	13	16	19	23	
	x_2:	–1	1	3	5	7	9	11	13	
	x_3:	3	6	9	12	15	18	21	24	
	Отклик объекта									
	y:	9.52	7.46	5.48	3.96	2.57	1.03	–0.85	–1.63	
15	Факторы									1,3
	x_1:	–14	–10	–7	–3	0	4	9	12	
	x_2:	–13	–11	–10	–6	–2	3	7	10	
	x_3:	–16	–14	–11	–7	–2	2	6	9	
	Отклик объекта									
	y:	–14.2	–8.74	–1.67	3.76	4.87	1.14	–2.39	–7.34	
16	Факторы									2,4
	x_1:	0	4	8	12	16	20	24	28	
	x_2:	0	5	10	15	20	25	30	35	
	x_3:	0	3	6	9	12	15	18	21	
	Отклик объекта									
	y:	–8.45	–6.49	–3.77	0.96	2.17	1.04	–0.59	–2.34	

Порядок выполнения работы

1. Сформировать функцию остаточной суммы от коэффициентов первого из указанных в задании полиномов регрессии.

2. Найти значения коэффициентов, при которых функция остаточной суммы достигает минимума.

3. Определить коэффициент корреляции первого полинома регрессии с таблицей данных эксперимента.

4. Выполнить действия п.п. 1–3 для второго полинома регрессии.

5. Определить значения остаточной суммы для первого и второго полиномов при оптимальных значениях их коэффициентов. Обосновать выбор одного из полиномов как более приемлемого.

Контрольные вопросы

1. Какую математическую модель называют идеальной математической моделью функции отклика объекта исследования?

2. Почему в качестве регрессионной математической модели объекта исследования чаще всего используется алгебраический степенной полином?

3. Из какой исходной функции формируется система альтернативных уравнений регрессии? Почему?

4. Как из уравнения регрессии конкретного вида формируется система условных уравнений?

5. Что такое "остаточная сумма" и "наилучшее решение" системы условных уравнений?

11.2 Лабораторная работа № 2. Предварительная обработка данных эксперимента, формирование и проверка адекватности уравнения регрессии

Задание: Реализовать методику предварительной обработки данных эксперимента, представленных в табл. 11.2, сформировать уравнение регрессии и определить степень его адекватности объекту исследования.

Таблица 11.2. Варианты заданий к лабораторной работе № 2

<table>
<tr><td>№ вар-та</td><td colspan="9">Таблицы данных эксперимента</td></tr>
<tr><td>1</td><td colspan="9">2</td></tr>
<tr><td rowspan="9">1</td><td colspan="9">Факторы эксперимента</td></tr>
<tr><td>x_1:</td><td>0</td><td>4</td><td>8</td><td>12</td><td>16</td><td>20</td><td>24</td><td>28</td></tr>
<tr><td>x_2:</td><td>–16</td><td>–14</td><td>–11</td><td>–7</td><td>–4</td><td>–2</td><td>1</td><td>3</td></tr>
<tr><td>x_3:</td><td>–10</td><td>–8</td><td>–6</td><td>–4</td><td>0</td><td>2</td><td>6</td><td>9</td></tr>
<tr><td>x_4:</td><td>–7</td><td>–5</td><td>–3</td><td>–1</td><td>1</td><td>3</td><td>5</td><td>7</td></tr>
<tr><td colspan="9">Отклик объекта исследования (3 параллельных опыта)</td></tr>
<tr><td rowspan="3">y:</td><td>14.2</td><td>8.74</td><td>1.67</td><td>–1.76</td><td>–3.07</td><td>–1.14</td><td>1.39</td><td>4.34</td></tr>
<tr><td>8.45</td><td>6.49</td><td>3.77</td><td>–0.96</td><td>–2.17</td><td>–1.04</td><td>0.59</td><td>2.34</td></tr>
<tr><td>9.52</td><td>7.46</td><td>2.48</td><td>0.85</td><td>–2.57</td><td>–1.03</td><td>0.79</td><td>3.63</td></tr>
</table>

Продолжение таблицы 11.2

1	*2*							
2	Факторы эксперимента							
	x_1:	0	3	7	9	13	16	19	23
	x_2:	–1	2	6	8	12	15	18	21
	x_3:	–18	–15	–12	–8	–5	–3	2	4
	x_4:	–11	–8	–5	0	2	5	8	11
	Отклик объекта исследования (3 параллельных опыта)								
	y:	6.25	4.08	3.85	3.44	0.81	–0.12	–3.11	–2.48
		8.04	5.12	4.32	3.86	1.31	–0.93	–4.42	–3.11
		7.67	4.38	4.05	3.62	1.04	–0.59	–3.87	–2.94
3	Факторы эксперимента								
	x_1:	0	5	10	15	20	25	30	35
	x_2:	–21	–18	–15	–11	–8	–5	–1	2
	x_3:	–19	–15	–9	–6	–3	–1	1	3
	x_4:	–9	–5	–2	0	3	6	9	11
	Отклик объекта исследования (3 параллельных опыта)								
	y:	–4.44	–2.86	0.54	1.99	2.43	4.21	5.63	6.26
		–6.38	–4.16	–0.32	1.05	3.27	5.08	6.35	7.78
		–5.17	–3.54	–0.17	1.65	3.03	4.98	5.86	7.32
4	Факторы эксперимента								
	x_1:	3	6	9	12	15	18	21	24
	x_2:	–17	–15	–12	–9	–5	–1	2	5
	x_3:	1	5	11	15	18	21	24	28
	x_4:	–11	–8	–5	–2	1	4	7	10
	Отклик объекта исследования (3 параллельных опыта)								
	y:	–2.87	–0.91	0.17	2.54	1.02	–1.14	–4.24	–6.96
		–4.65	–2.48	–0.72	1.86	2.35	–1.39	–3.42	–5.82
		–3.69	–1.86	–0.33	1.42	2.08	–1.28	–4.81	–6.37
5	Факторы эксперимента								
	x_1:	1	4	10	16	19	23	26	30
	x_2:	–10	–6	–3	0	2	6	9	11
	x_3:	–21	–18	–15	–12	–9	–6	–3	0
	x_4:	–11	–8	–5	–2	1	4	7	10
	Отклик объекта исследования (3 параллельных опыта)								
	y:	–3.71	–2.18	–0.54	5.34	12.87	25.71	42.95	78.63
		–2.25	–1.11	0.86	3.62	10.49	22.32	37.41	80.42
		–4.37	–2.92	–0.25	4.34	11.71	24.02	40.85	81.86

Продолжение таблицы 11.2

1	*2*								
6	Факторы эксперимента								
	x_1:	2	5	8	13	18	22	25	28
	x_2:	–9	–6	–3	0	3	6	9	12
	x_3:	–23	–19	–15	–11	–7	–4	0	2
	x_4:	0	3	9	14	20	24	27	29
	Отклик объекта исследования (3 параллельных опыта)								
	y:	81.23	44.23	16.47	9.4	4.14	1.47	–0.28	–1.35
		78.32	43.64	15.26	8.29	3.45	1.19	–1.05	–2.16
		83.05	46.03	17.74	10.43	5.24	2.21	–1.82	–3.08
7	Факторы эксперимента								
	x_1:	–2	3	9	13	18	23	27	32
	x_2:	–26	–22	–18	–12	–8	–4	0	4
	x_3:	–9	–6	–3	0	3	6	9	12
	x_4:	–13	–10	–7	–3	2	3	7	10
	Отклик объекта исследования (3 параллельных опыта)								
	y:	5.27	3.06	–0.85	–1.33	0.41	0.92	–1.41	–2.37
		6.04	2.72	–0.32	–1.56	0.71	1.13	–1.28	–3.64
		4.87	2.38	–0.15	–0.82	1.04	0.79	–0.87	–2.71
8	Факторы эксперимента								
	x_1:	–1	4	8	10	15	19	24	29
	x_2:	–2	3	7	11	14	18	23	27
	x_3:	–25	–21	–17	–13	–9	–5	–1	3
	x_4:	–12	–9	–7	–4	0	4	8	11
	Отклик объекта исследования (3 параллельных опыта)								
	y:	–4.44	–2.86	0.54	1.99	–1.43	–0.96	1.63	3.26
		–6.13	–4.62	1.04	2.63	–0.73	–1.43	0.88	2.48
		–5.58	–3.71	1.13	2.38	–1.21	–0.77	1.29	2.85
9	Факторы эксперимента								
	x_1:	0	4	8	12	16	20	24	28
	x_2:	–16	–14	–11	–7	–4	–2	1	3
	x_3:	–10	–8	–6	–4	0	2	6	9
	x_4:	–7	–5	–3	–1	1	3	5	7
	Отклик объекта исследования (3 параллельных опыта)								
	y:	–14.2	–8.74	–1.67	1.76	3.07	1.14	–1.39	–4.34
		–8.45	–6.49	–3.77	0.96	2.17	1.04	–0.59	–2.34
		–9.52	–7.46	–2.48	–0.85	2.57	1.03	–0.79	–3.63

Продолжение таблицы 11.2

1	*2*								
10	Факторы эксперимента								
	x_1:	0	3	7	9	13	16	19	23
	x_2:	–1	2	6	8	12	15	18	21
	x_3:	–18	–15	–12	–8	–5	–3	2	4
	x_4:	–11	–8	–5	0	2	5	8	11
	Отклик объекта исследования (3 параллельных опыта)								
	y:	–6.25	–4.08	–3.85	–3.44	–0.81	0.12	3.11	2.48
		–8.04	–5.12	–4.32	–3.86	–1.31	0.93	4.42	3.11
		–7.67	–4.38	–4.05	–3.62	–1.04	0.59	3.87	2.94
11	Факторы эксперимента								
	x_1:	0	5	10	15	20	25	30	35
	x_2:	–21	–18	–15	–11	–8	–5	–1	2
	x_3:	–19	–15	–9	–6	–3	–1	1	3
	x_4:	–9	–5	–2	0	3	6	9	11
	Отклик объекта исследования (3 параллельных опыта)								
	y:	4.44	2.86	–0.54	–1.99	–2.43	–4.21	–5.63	–6.26
		6.38	4.16	0.32	–1.05	–3.27	–5.08	–6.35	–7.78
		5.17	3.54	0.17	–1.65	–3.03	–4.98	–5.86	–7.32
12	Факторы эксперимента								
	x_1:	3	6	9	12	15	18	21	24
	x_2:	–17	–15	–12	–9	–5	–1	2	5
	x_3:	1	5	11	15	18	21	24	28
	x_4:	–11	–8	–5	–2	1	4	7	10
	Отклик объекта исследования (3 параллельных опыта)								
	y:	2.87	0.91	–0.17	–2.54	–1.02	1.14	4.24	6.96
		4.65	2.48	0.72	–1.86	–2.35	1.39	3.42	5.82
		3.69	1.86	0.33	–1.42	–2.08	1.28	4.81	6.37
13	Факторы эксперимента								
	x_1:	1	4	10	16	19	23	26	30
	x_2:	–10	–6	–3	0	2	6	9	11
	x_3:	–21	–18	–15	–12	–9	–6	–3	0
	x_4:	–11	–8	–5	–2	1	4	7	10
	Отклик объекта исследования (3 параллельных опыта)								
	y:	3.71	2.18	0.54	–5.34	–12.87	–25.71	–42.95	–78.63
		2.25	1.11	–0.86	–3.62	–10.49	–22.32	–37.41	–80.42
		4.37	2.92	0.25	–4.34	–11.71	–24.02	–40.85	–81.86

Продолжение таблицы 11.2

1	2								
14	Факторы эксперимента								
	x_1:	2	5	8	13	18	22	25	28
	x_2:	–9	–6	–3	0	3	6	9	12
	x_3:	–23	–19	–15	–11	–7	–4	0	2
	x_4:	0	3	9	14	20	24	27	29
	Отклик объекта исследования (3 параллельных опыта)								
	y:	–81.23	–44.23	–16.47	–9.4	4.14	–1.47	0.28	1.35
		–78.32	–43.64	–15.26	–8.29	3.45	–1.19	1.05	2.16
		–83.05	–46.03	–17.74	–10.43	5.24	–2.21	1.82	3.08
15	Факторы эксперимента								
	x_1:	–2	3	9	13	18	23	27	32
	x_2:	–26	–22	–18	–12	–8	–4	0	4
	x_3:	–9	–6	–3	0	3	6	9	12
	x_4:	–13	–10	–7	–3	2	3	7	10
	Отклик объекта исследования (3 параллельных опыта)								
	y:	–5.27	–3.06	0.85	1.33	–0.41	–0.92	1.41	2.37
		–6.04	–2.72	0.32	1.56	–0.71	–1.13	1.28	3.64
		–4.87	–2.38	0.15	0.82	–1.04	–0.79	0.87	2.71
16	Факторы эксперимента								
	x_1:	–1	4	8	10	15	19	24	29
	x_2:	–2	3	7	11	14	18	23	27
	x_3:	–25	–21	–17	–13	–9	–5	–1	3
	x_4:	–12	–9	–7	–4	0	4	8	11
	Отклик объекта исследования (3 параллельных опыта)								
	y:	4.44	2.86	–0.54	–1.99	1.43	0.96	–1.63	–3.26
		6.13	4.62	–1.04	–2.63	0.73	1.43	–0.88	–2.48
		5.58	3.71	–1.13	–2.38	1.21	0.77	–1.29	–2.85

Порядок выполнения работы

1. Определить коэффициент корреляции каждого из предложенных факторов эксперимента с каждым и принять решение о возможности отказа от использования какого-либо из факторов.

2. Выявить ошибочные опыты с помощью критерия Стьюдента при уровне значимости 0.05, см. приложение А, и удалить соответствующие элементы из таблицы данных эксперимента.

3. Проверить условие воспроизводимости опытов по критерию Кохрена при уровне значимости 0.05, см. приложение Б, и, при его невыполнении, указать на результаты опытов, вызывающие сомнения в их достоверности.

4. Выбрать вид уравнения регрессии и найти значения его коэффициентов, при которых функция остаточной суммы достигает минимума. Проверить значимость коэффициентов по критерию Стьюдента. Удалить слагаемые с незначимы-

ми коэффициентами из уравнения регрессии и вновь определить значения его коэффициентов.

5. Определить расчетные значения отклика объекта исследования и значение корреляционного отношения между ними и усредненными значениями результатов экспериментов.

6. Определить степень адекватности сформированного уравнения регрессии объекту исследования по критерию Фишера при уровне значимости 0.05, см. приложение В.

Контрольные вопросы

1. Что является показателем ошибочности результатов данного конкретного эксперимента?
2. Какие действия рекомендуется предпринять при невыполнении условия воспроизводимости опытов?
3. Какую операцию необходимо реализовать после удаления из уравнения регрессии слагаемых, коэффициенты при которых можно считать незначимыми?
4. Каков физический смысл критерия Фишера?
5. В чем суть операции интерпретации полинома регрессии, реализуемой после установления факта его адекватности объекту исследования?

11.3 Лабораторная работа № 3. Полный/дробный факторный эксперимент и обработка его результатов

Задание: Сформировать математическую модель влияния указанных в табл. 11.3 технологических факторов на качество поверхности магнитного диска (ее максимальную неровность, мкм) с применением методики полного (ПФЭ) или дробного (ДФЭ) факторного эксперимента. Найти натуральные значения факторов, которым соответствует условный минимум полинома регрессии. Реализовать процедуру интерпретации полинома регрессии.

Данные экспериментов сведены в табл. 10.4. Если вариантом задания предусмотрен ДФЭ, то в графе "Вид эксперимента, факторы" этой таблицы указываются только генераторы плана для факторов, которые отнесены к второстепенным.

Таблица 11.3 Наименования факторов и уровни их варьирования

Факторы эксперимента	Уровни варьирования		
	–1	0	+1
Опорное напряжение – x_1, В	28	30	32
Ток потребления – x_2, А	17	18	19
Конечная температура нагрева – x_3, ^{o}С	200	220	240
Скорость нагрева – x_4, ^{o}С/с	7	10	13
Продолжительность изотермической выдержки – x_5, с	65	80	95

Таблица 11.4 Вид, факторы и результаты экспериментов

№ вар-та	Вид эксп-та, факторы	От-кли-ки	Максимальная неровность поверхности диска, мкм							
			№ эксперимента (3 параллельных опыта)							
			1	2	3	4	5	6	7	8
1	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*
1	ПФЭ, x_1, x_2, x_3	y_1	7.87	17.23	6.55	8.49	20.16	32.14	6.85	27.19
		y_2	7.41	15.42	5.89	10.91	19.84	27.59	7.2	23.56
		y_3	11.2	15.64	11.26	8.79	22.39	28.16	9.64	23.04
2	ДФЭ, $x_4 = x_1x_2$; $x_5 = x_1x_2x_3$	y_1	6.55	20.16	6.85	17.23	8.49	27.19	7.87	32.14
		y_2	5.89	19.84	7.2	15.42	10.91	23.56	7.41	27.59
		y_3	11.26	22.39	9.64	15.64	8.79	23.04	11.2	28.16
3	ПФЭ, x_1, x_2, x_4	y_1	6.37	4.0	2.96	1.16	5.06	2.74	2.96	2.46
		y_2	6.19	3.59	3.96	0.86	4.87	2.94	2.44	2.14
		y_3	6.27	3.87	3.75	1.82	4.87	2.61	2.8	2.8
4	ДФЭ, $x_3 = x_1x_4$; $x_5 = x_1x_2x_4$	y_1	2.96	1.16	6.37	4.0	2.96	2.46	5.06	2.74
		y_2	3.96	0.86	6.19	3.59	2.44	2.14	4.87	2.94
		y_3	3.75	1.82	6.27	3.87	2.8	2.8	4.87	2.61
5	ПФЭ, x_1, x_2, x_5	y_1	2.53	7.37	0.6	8.47	4.68	6.19	0.35	3.66
		y_2	3.9	7.13	0.44	6.88	3.7	5.32	0.19	3.4
		y_3	3.26	7.95	1.12	7.44	3.22	6.76	0.75	4.23
6	ДФЭ, $x_3 = x_1x_2$; $x_4 = x_1x_2x_5$	y_1	0.35	3.66	4.68	6.19	0.6	8.47	2.53	7.37
		y_2	0.19	3.4	3.7	5.32	0.44	6.88	3.9	7.13
		y_3	0.75	4.23	3.22	6.76	1.12	7.44	3.26	7.95
7	ПФЭ, x_1, x_3, x_4	y_1	9.7	3.7	7.8	4.2	5.3	4.5	1.8	4.3
		y_2	10.4	3.4	7.5	4.4	5.7	5.1	2.5	4.2
		y_3	11.4	4.1	7.7	4.5	6.2	4.8	2	4.9
8	ДФЭ, $x_2 = x_3x_4$; $x_5 = x_1x_3x_4$	y_1	4.2	7.8	3.7	9.7	4.3	1.8	4.5	5.3
		y_2	4.4	7.5	3.4	10.4	4.2	2.5	5.1	5.7
		y_3	4.5	7.7	4.1	11.4	4.9	2	4.8	6.2
9	ПФЭ, x_1, x_3, x_5	y_1	7.87	17.23	6.55	8.49	20.16	32.14	6.85	27.19
		y_2	7.41	15.42	5.89	10.91	19.84	27.59	7.2	23.56
		y_3	11.2	15.64	11.26	8.79	22.39	28.16	9.64	23.04
10	ДФЭ, $x_2 = x_1x_3$; $x_4 = x_1x_3x_5$	y_1	6.55	20.16	6.85	17.23	8.49	27.19	7.87	32.14
		y_2	5.89	19.84	7.2	15.42	10.91	23.56	7.41	27.59
		y_3	11.26	22.39	9.64	15.64	8.79	23.04	11.2	28.16
11	ПФЭ, x_1, x_4, x_5	y_1	6.37	4.0	2.96	1.16	5.06	2.74	2.96	2.46
		y_2	6.19	3.59	3.96	0.86	4.87	2.94	2.44	2.14
		y_3	6.27	3.87	3.75	1.82	4.87	2.61	2.8	2.8
12	ДФЭ, $x_2 = x_1x_4$; $x_3 = x_1x_4x_5$	y_1	2.96	1.16	6.37	4.0	2.96	2.46	5.06	2.74
		y_2	3.96	0.86	6.19	3.59	2.44	2.14	4.87	2.94
		y_3	3.75	1.82	6.27	3.87	2.8	2.8	4.87	2.61
13	ПФЭ, x_2, x_3, x_4	y_1	2.53	7.37	0.6	8.47	4.68	6.19	0.35	3.66
		y_2	3.9	7.13	0.44	6.88	3.7	5.32	0.19	3.4
		y_3	3.26	7.95	1.12	7.44	3.22	6.76	0.75	4.23

Продолжение таблицы 11.4

1	2	3	4	5	6	7	8	9	10	11
14	ДФЭ, $x_1 = x_2x_3$; $x_5 = x_2x_3x_4$	y_1	0.35	3.66	4.68	6.19	0.6	8.47	2.53	7.37
		y_2	0.19	3.4	3.7	5.32	0.44	6.88	3.9	7.13
		y_3	0.75	4.23	3.22	6.76	1.12	7.44	3.26	7.95
15	ПФЭ, x_2, x_3, x_5	y_1	9.7	3.7	7.8	4.2	5.3	4.5	1.8	4.3
		y_2	10.4	3.4	7.5	4.4	5.7	5.1	2.5	4.2
		y_3	11.4	4.1	7.7	4.5	6.2	4.8	2	4.9
16	ДФЭ, $x_1 = x_3x_5$; $x_4 = x_2x_3x_5$	y_1	4.2	7.8	3.7	9.7	4.3	1.8	4.5	5.3
		y_2	4.4	7.5	3.4	10.4	4.2	2.5	5.1	5.7
		y_3	4.5	7.7	4.1	11.4	4.9	2	4.8	6.2

Порядок выполнения работы

1. Представить результаты эксперимента в виде совокупности векторов. В качестве значений факторов для каждого из опытов использовать –1 или 1. Если вариантом задания предусмотрен ДФЭ, определить контрасты вспомогательных факторов и определяющий контраст, смешанность оценок коэффициентов уравнения регрессии.

2. Реализовать процедуру проверки ошибочности опытов с использованием критерия Стьюдента при уровне значимости 0.05, см. приложение А.

3. Проверить выполнение условия воспроизводимости опытов с применением критерия Кохрена при уровне значимости 0.05, см. приложение Б.

4. Рассчитать значения коэффициентов уравнения регрессии и проверить их значимость с использованием критерия Стьюдента при уровне значимости 0.05. Сформировать уравнение регрессии из тех слагаемых, которым соответствуют значимые коэффициенты.

5. Определить расчетные значения отклика объекта исследования и проверить выполнение условия адекватности уравнения регрессии с использованием критерия Фишера при уровне значимости 0.05, см. приложение В.

6. Определить условный минимум полинома регрессии и оптимальные значения факторов в натуральном выражении.

7. Реализовать операцию интерпретации полинома регрессии, установить степень влияния кодированных значений факторов на значение отклика объекта исследования.

Контрольные вопросы

1. Как кодируются значения факторов при планировании эксперимента?

2. Как влияет свойство ортогональности матриц планирования ПФЭ и ДФЭ на методику оценки значимости коэффициентов уравнения регрессии.

3. Что предопределяет свойство ротатабельности матрицы планирования ПФЭ и ДФЭ?

4. Укажите преимущество ДФЭ перед ПФЭ.

5. Какая априорная информация необходима для выбора конкретной дробной реплики ДФЭ?

11.4 Лабораторная работа №4. Центральные композиционные планы эксперимента и обработка результатов их применения

<u>Задание</u>: сформировать математическую модель влияния трех технологических факторов на силу резания древесного сырья при цилиндрическом фрезеровании с применением центрального композиционного плана эксперимента указанного вида, найти натуральные значения факторов, которым соответствует условный минимум полинома регрессии. Наименования факторов и уровни их варьирования приведены в табл. 11.5, – результаты экспериментов – в табл. 11.6.

Виды планов эксперимента: ортогональный (ОЦКП) – варианты 1– 4;
ортогональный вписанный (ВОЦКП) – варианты 5–8;
гранецентрированный (ГЦКП) – варианты 9–12;
ротатабельный (РЦКП) – варианты 13–16.

Таблица 11.5. Наименования факторов и уровни их варьирования

Факторы	Уровни варьирования				
	–α*	–g**	0	+g**	+α*
Ширина фрезерования – x_1, мм	78.5	100	200	300	321.5
Время фрезерования – x_2, мин.	34	60	180	300	326
Скорость подачи сырья – x_3, м/мин.	6.28	8.0	16.0	24.0	25.72

$$* \ \alpha = \begin{cases} 1.215 - \text{ОЦКП}, \\ 1 - \text{ВОЦКП}, \\ \text{не используется} - \text{ГЦКП}, \\ 1.682 - \text{РЦКП}. \end{cases} \qquad ** \ g = \begin{cases} 1 - \text{ОЦКП, ГЦКП, РЦКП}, \\ 0.823 - \text{ВОЦКП}. \end{cases}$$

Таблица 11.6. Результаты экспериментов

№ вар-та	От-клик	Сила резания, Н														
		№ эксперимента (три серии опытов)														
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*	*11*	*12*	*13*	*14*	*15*	*16*	*17*
1	y_1	17.67	51.74	18.9	78.58	24.69	97.15	23.86	142.63	16.93	86.03	47.08	62.63	38.59	81.2	63.71
	y_2	2.15	50.72	23.15	79.01	41.0	92.37	43.05	114.66	20.13	80.61	52.19	55.7	39.76	93.72	64.79
	y_3	5.51	86.2	6.57	56.19	4.13	95.44	32.28	109.06	26.4	119.14	54.49	63.37	56.46	88.55	66.93
2	y_1	7.78	48.49	43.64	82.28	61.82	96.01	50.64	160.25	13.28	118.74	30.41	81.46	52.6	95.48	71.36
	y_2	36.56	46.14	17.48	57.57	40.53	85.38	39.76	129.01	29.81	122.1	50.37	88.51	28.08	76.7	66.88
	y_3	40.85	44.96	34.32	76.02	24.49	122.44	67.85	119.49	45.05	100.46	42.23	84.11	39.43	53.15	54.72
3	y_1	17.67	51.74	18.9	78.58	24.69	97.15	23.86	142.63	16.93	86.03	47.08	62.63	38.59	81.2	63.71
	y_2	2.15	50.72	23.15	79.01	41.0	92.37	43.05	114.66	20.13	80.61	52.19	55.7	39.76	93.72	64.79
	y_3	7.78	48.49	43.64	82.28	61.82	96.01	50.64	160.25	13.28	118.74	30.41	81.46	52.6	95.48	71.36
4	y_1	17.67	51.74	18.9	78.58	24.69	97.15	23.86	142.63	16.93	86.03	47.08	62.63	38.59	81.2	63.71
	y_2	5.51	86.2	6.57	56.19	4.13	95.44	32.28	109.06	26.4	119.14	54.49	63.37	56.46	88.55	66.93
	y_3	7.78	48.49	43.64	82.28	61.82	96.01	50.64	160.25	13.28	118.74	30.41	81.46	52.6	95.48	71.36
5	y_1	2.15	50.72	23.15	79.01	41.0	92.37	43.05	114.66	20.13	80.61	52.19	55.7	39.76	93.72	64.79
	y_2	5.51	86.2	6.57	56.19	4.13	95.44	32.28	109.06	26.4	119.14	54.49	63.37	56.46	88.55	66.93
	y_3	7.78	48.49	43.64	82.28	61.82	96.01	50.64	160.25	13.28	118.74	30.41	81.46	52.6	95.48	71.36

Продолжение таблицы 11.6

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17
6	y_1	17.67	51.74	18.9	78.58	24.69	97.15	23.86	142.63	16.93	86.03	47.08	62.63	38.59	81.2	63.71
	y_2	2.15	50.72	23.15	79.01	41.0	92.37	43.05	114.66	20.13	80.61	52.19	55.7	39.76	93.72	64.79
	y_3	36.56	46.14	17.48	57.57	40.53	85.38	39.76	129.01	29.81	122.1	50.37	88.51	28.08	76.7	66.88
7	y_1	2.15	50.72	23.15	79.01	41.0	92.37	43.05	114.66	20.13	80.61	52.19	55.7	39.76	93.72	64.79
	y_2	5.51	86.2	6.57	56.19	4.13	95.44	32.28	109.06	26.4	119.14	54.49	63.37	56.46	88.55	66.93
	y_3	36.56	46.14	17.48	57.57	40.53	85.38	39.76	129.01	29.81	122.1	50.37	88.51	28.08	76.7	66.88
8	y_1	17.67	51.74	18.9	78.58	24.69	97.15	23.86	142.63	16.93	86.03	47.08	62.63	38.59	81.2	63.71
	y_2	5.51	86.2	6.57	56.19	4.13	95.44	32.28	109.06	26.4	119.14	54.49	63.37	56.46	88.55	66.93
	y_3	36.56	46.14	17.48	57.57	40.53	85.38	39.76	129.01	29.81	122.1	50.37	88.51	28.08	76.7	66.88
9	y_1	7.78	48.49	43.64	82.28	61.82	96.01	50.64	160.25	13.28	118.74	30.41	81.46	52.6	95.48	71.36
	y_2	36.56	46.14	17.48	57.57	40.53	85.38	39.76	129.01	29.81	122.1	50.37	88.51	28.08	76.7	66.88
	y_3	17.67	51.74	18.9	78.58	24.69	97.15	23.86	142.63	16.93	86.03	47.08	62.63	38.59	81.2	63.71
10	y_1	17.67	51.74	18.9	78.58	24.69	97.15	23.86	142.63	16.93	86.03	47.08	62.63	38.59	81.2	63.71
	y_2	2.15	50.72	23.15	79.01	41.0	92.37	43.05	114.66	20.13	80.61	52.19	55.7	39.76	93.72	64.79
	y_3	36.56	46.14	17.48	57.57	40.53	85.38	39.76	129.01	29.81	122.1	50.37	88.51	28.08	76.7	66.88
11	y_1	2.15	50.72	23.15	79.01	41.0	92.37	43.05	114.66	20.13	80.61	52.19	55.7	39.76	93.72	64.79
	y_2	5.51	86.2	6.57	56.19	4.13	95.44	32.28	109.06	26.4	119.14	54.49	63.37	56.46	88.55	66.93
	y_3	40.85	44.96	34.32	76.02	24.49	122.44	67.85	119.49	45.05	100.46	42.23	84.11	39.43	53.15	54.72
12	y_1	17.67	51.74	18.9	78.58	24.69	97.15	23.86	142.63	16.93	86.03	47.08	62.63	38.59	81.2	63.71
	y_2	5.51	86.2	6.57	56.19	4.13	95.44	32.28	109.06	26.4	119.14	54.49	63.37	56.46	88.55	66.93
	y_3	40.85	44.96	34.32	76.02	24.49	122.44	67.85	119.49	45.05	100.46	42.23	84.11	39.43	53.15	54.72
13*	y_1	7.78	48.49	43.64	82.28	61.82	96.01	50.64	160.25	13.28	118.74	30.41	81.46	52.6	95.48	71.36
	y_2	40.85	44.96	34.32	76.02	24.49	122.44	67.85	119.49	45.05	100.46	42.23	84.11	39.43	53.15	54.72
	y_3	17.67	51.74	18.9	78.58	24.69	97.15	23.86	142.63	16.93	86.03	47.08	62.63	38.59	81.2	63.71
14*	y_1	36.56	46.14	17.48	57.57	40.53	85.38	39.76	129.01	29.81	122.1	50.37	88.51	28.08	76.7	66.88
	y_2	40.85	44.96	34.32	76.02	24.49	122.44	67.85	119.49	45.05	100.46	42.23	84.11	39.43	53.15	54.72
	y_3	17.67	51.74	18.9	78.58	24.69	97.15	23.86	142.63	16.93	86.03	47.08	62.63	38.59	81.2	63.71
15*	y_1	7.78	48.49	43.64	82.28	61.82	96.01	50.64	160.25	13.28	118.74	30.41	81.46	52.6	95.48	71.36
	y_2	36.56	46.14	17.48	57.57	40.53	85.38	39.76	129.01	29.81	122.1	50.37	88.51	28.08	76.7	66.88
	y_3	2.15	50.72	23.15	79.01	41.0	92.37	43.05	114.66	20.13	80.61	52.19	55.7	39.76	93.72	64.79
16*	y_1	7.78	48.49	43.64	82.28	61.82	96.01	50.64	160.25	13.28	118.74	30.41	81.46	52.6	95.48	71.36
	y_2	40.85	44.96	34.32	76.02	24.49	122.44	67.85	119.49	45.05	100.46	42.23	84.11	39.43	53.15	54.72
	y_3	2.15	50.72	23.15	79.01	41.0	92.37	43.05	114.66	20.13	80.61	52.19	55.7	39.76	93.72	64.79

* результаты опытов №№ 16–20 принять близкими к результатам опыта № 15 (отклонение в пределах 5%).

Порядок выполнения работы

1. Сформировать столбцы кодированных значений факторов согласно матрице ЦКП, указанного в задании.

2. Определить усредненные значения откликов и проверить ошибочность опытов по критерию Стьюдента и воспроизводимость опытов по критерию Кохрена.

3. Оценить дисперсию воспроизводимости опытов, рассчитать значения коэффициентов полинома регрессии. Определить среднеквадратичное отклонение коэффициентов полинома регрессии и выбрать незначимые коэффициенты.

4. Для ГЦКП и РЦКП вновь определить значимые коэффициенты полинома регрессии.

5. Записать полином регрессии со значимыми коэффициентами и определить расчетные значения отклика объекта. Рассчитать остаточную дисперсию и оценить адекватность полинома регрессии отклику объекта исследования.

6. Найти кодированные значения факторов, которым соответствует условный минимум отклика объекта. Перейти от кодированных значений факторов к натуральным.

Контрольные вопросы

1. Укажите преимущество центрального композиционного планирования эксперимента перед полным/дробным факторным экспериментом.

2. Какой из центральных композиционных планов эксперимента может быть использован для построения полинома регрессии третьего порядка?

3. Укажите отличия и преимущества ВОЦКП и ГЦКП перед ОЦКП, преимущество ОЦКП и ВОЦКП перед ГЦКП.

4. Дайте определение свойства ротатабельности ЦКП и укажите преимущество и недостаток РЦКП по сравнению с ОЦКП.

5. Почему после определения экспериментальной области факторного пространства рекомендуется реализовать процедуру поиска экстремума поверхности отклика объекта?

ЗАКЛЮЧЕНИЕ

Предлагаемое пособие включает основные сведения о методике планирования эксперимента и обработке его результатов с целью построения регрессионной математической модели объекта исследования.

В пособии рассмотрены:

- рекомендации по выбору факторов эксперимента, откликов объекта исследования;
- виды уравнений регрессии и условия выбора вида уравнения в конкретных ситуациях;
- основные понятия математической статистики и стохастической связи между случайными величинами;
- методика предварительной обработки результатов экспериментов;
- последовательность формирования и проверки адекватности уравнения регрессии объекту исследования;
- методика проведения и обработки результатов полного и дробного факторного эксперимента;
- разновидности центральных композиционных планов эксперимента и методика обработки результатов их применения.

Лабораторный практикум содержит формулировки заданий, исходные данные, рекомендации по выполнению и контрольные вопросы к четырем лабораторным работам:

- формирование полиномов регрессии заданной конфигурации и оценка степени их корреляции с таблицей данных эксперимента;
- предварительная обработка данных эксперимента, формирование и проверка адекватности уравнения регрессии;
- полный/дробный факторный эксперимент и обработка его результатов;
- центральные композиционные планы эксперимента и обработка результатов их применения.

Авторы надеются, что предлагаемое пособие будет полезно для студентов, обучающихся по направлениям 15.03.01 и 15.04.01 – "Машиностроение", 15.04.05 – "Конструкторско-технологическое обеспечение машиностроительных производств" дневной и заочной формы обучения, а также для аспирантов и работников проектно-конструкторских отделов промышленных предприятий.

ЛИТЕРАТУРА

1. Пучков, Н.П. Математическая статистика. Применение в профессиональной деятельности: учебное пособие / Н.П. Пучков. – Тамбов: Изд-во ФГБОУ ВПО "ТГТУ", 2013. – 80 с.

2. Федоров, В.В. Теория оптимального эксперимента / В.В. Федоров. – М.: Наука, 1971. – 312 с.

3. Адлер, Ю.П. Планирование эксперимента при поиске оптимальных условий / Ю.П. Адлер, Е.В. Маркова, Ю.В. Грановский. – М.: Наука, 1976. – 280 с.

4. Джонсон, И. Статистика и планирование эксперимента в технике и науке. Методы планирования эксперимента. Пер. с англ. / И. Джонсон, Ф. Лион. – М. : Мир, 1981. – 520 с.

5. Бородюк, В.П. Статистические методы в инженерных исследованиях / В.П. Бородюк, А.П. Вощинин, А.З. Иванов и др. – М.: Высшая школа,1983. – 216 с.

6. Ахназарова, С.Л. Методы оптимизации эксперимента в химической технологии: Учеб. пособие для хим.-технол. спец. вузов / С.Л. Ахназарова, В.В. Кафаров. – М.: Высш. шк., 1985. – 327 с.

7. Современный эксперимент : подготовка, проведение, анализ результатов : учебник для вузов / В. Г. Блохин, О. П. Глудких, А. И. Гуров, Н. А. Ханин; под ред. О. П. Глудких. – М. : Радио и связь, 1997. – 326 с.

8. Рогов, В. А. Методика и практика технических экспериментов : учебное пособие для вузов / В. А. Рогов, Г. Г. Позняк. – М. : Академия, 2005. – 283 с.

9. Сидняев, Н.И. Теория планирования эксперимента и анализ статистических данных : учебное пособие / Н.И. Сидняев. – М. : ИД Юрайт, 2012. – 399 с.

10. Нинул, А.С. Оптимизация целевых функций: Аналитика. Численные методы. Планирование эксперимента / А.С. Нинул. – М.: Издательство Физико-математической литературы, 2009. – 336 с.

11. Спиридонов, А.А. Планирование эксперимента при исследовании технологических процессов / А.А. Спиридонов. М.: Машиностроение, 1981. – 184 с.

12. Box, G.E.P. Some New Three Level Designs for the Study of Quantitative Variables / G.E.P. Box, D.W. Behnken // Technometrics. 1960. Vol. 2. №4. P. 455-475.

Приложение А. Значения критерия Стьюдента

Значения критерия Стьюдента t для уровня значимости 0.05 и числа степеней свободы f:

f	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
t	**12.71**	**4.3**	**3.18**	**2.78**	**2.57**	**2.45**	**2.36**	**2.31**	**2.26**	**2.23**	**2.2**	**2.18**	**2.16**	**2.15**	**2.13**	**2.12**	**2.11**	**2.1**	**2.09**
f	20	21	22	23	24	25	26	27	28	29	30	40	50	60	80	100	120	200	≥500
t	**2.09**	**2.08**	**2.07**	**2.07**	**2.06**	**2.06**	**2.06**	**2.05**	**2.05**	**2.04**	**2.04**	**2.02**	**2.01**	**2**	**1.99**	**1.98**	**1.98**	**1.97**	**1.96**

Приложение Б. Значения критерия Кохрена

Значения критерия Кохрена при уровне значимости 0.05 для объема выборки N и числа степеней свободы f

N	f													
	1	2	3	4	5	6	7	8	9	10	16	36	144	∞
2	0.99	0.98	0.94	0.91	0.88	0.85	0.83	0.82	0,80	0.79	0.73	0.66	0.58	0.50
3	0.97	0.87	0.80	0.75	0.71	0.68	0.65	0.63	0.62	0.60	0.55	0.47	0.40	0.33
4	0.91	0.77	0.68	0.63	0.59	0.56	0.54	0.52	0.50	0.49	0.44	0.37	0.31	0.25
5	0.84	0.68	0.60	0.54	0.51	0.48	0.46	0.44	0.42	0.41	0.36	0.31	0.25	0.20
6	0.78	0.62	0.53	0.48	0.44	0.42	0.40	0.38	0.37	0.36	0.31	0.26	0.21	0.17
7	0.73	0.56	0.48	0.43	0.40	0.37	0.35	0.34	0.33	0.32	0.28	0.23	0.18	0.14
8	0.68	0.52	0.44	0.39	0.36	0.34	0.32	0.30	0.29	0.28	0.25	0.20	0.16	0.13
9	0.64	0.48	0.40	0.36	0.33	0.31	0.29	0.28	0.27	0.26	0.22	0.18	0.14	0.11
10	0.60	0.45	0.37	0.33	0.30	0.28	0.27	0.25	0.24	0.24	0.20	0.17	0.13	0.10
12	0.54	0.39	0.33	0.29	0.26	0.24	0.23	0.22	0.21	0.20	0.17	0.14	0.11	0.08
15	0.47	0.33	0.28	0.24	0.22	0.20	0.19	0.18	0.17	0017	0.14	0.11	0.09	0.07
20	0.39	0.27	0.22	0.19	0.17	0.16	0.15	0.14	0.14	0.13	0.11	0.09	0.07	0.05
24	0.34	0.24	0.19	0.17	0.15	0.14	0.13	0.12	0.12	0.11	0.09	0.07	0.06	0.04
30	0.29	0.20	0.16	0.14	0.12	0.11	0.11	0.10	0.10	0.09	0.08	0.06	0.05	0.03
40	0.24	0.16	0.13	0.11	0.10	0.09	0.08	0.08	0.07	0.07	0.06	0.05	0.03	0.03
60	0.17	0.11	0.09	0.09	0.07	0.06	0.06	0.06	0.05	0.05	0.04	0.03	0.02	0.02
120	0.10	0.06	0.05	0.04	0.04	0.03	0.03	0.03	0.03	0.03	0.02	0.02	0.01	0.01

Приложение В. Значения критерия Фишера

Значения критерия Фишера для уровня значимости 0.05 при числах степеней свободы f1 (для остаточной дисперсии) и f2 (для дисперсии воспроизводимости):

f_2	f_1														
	1	2	3	4	5	6	7	8	9	10	12	15	20	30	∞
1	161.0	200.0	216.0	225.0	230.0	234.0	237.0	239.0	241.0	242.0	244.0	246.0	248.0	250.0	254.0
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38	19.40	19.41	19.43	19.45	19.46	19.50
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79	8.74	8.70	8.66	8.62	8.53
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.94	5.91	5.86	5.80	5.75	5.63
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.68	4.62	4.56	4.50	4.36
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	4.00	3.94	3.87	3.81	3.67
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64	3.57	3.51	3.44	3.38	3.23
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35	3.28	3.22	3.15	3.08	2.93
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.91	2.85	2.77	2.70	2.54
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75	2.69	2.62	2.54	2.47	2.30
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60	2.53	2.46	2.39	2.31	2.13
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49	2.42	2.35	2.28	2.19	2.01
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41	2.34	2.27	2.19	2.11	1.92
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35	2.28	2.20	2.12	2.04	1.84
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34	2.30	2.23	2.15	2.07	1.98	1.78
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30	2.25	2.18	2.11	2.03	1.94	1.73
26	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27	2.22	2.15	2.07	1.99	1.90	1.69
28	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24	2.19	2.12	2.04	1.96	1.87	1.65
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16	2.09	2.01	1.93	1.84	1.62
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.08	2.00	1.92	1.84	1.74	1.51
60	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04	1.99	1.92	1.84	1.75	1.65	1.39
120	3.92	3.07	2.68	2.45	2.29	2.17	2.09	2.02	1.96	1.91	1.83	1.75	1.66	1.55	1.25
∞	3.84	3.00	2.60	2.37	2.21	2.10	2.01	1.94	1.88	1.83	1.75	1.67	1.57	1.46	1.00