МИНИСТЕРСТВО

ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ

Государственное образовательное учреждение высшего профессионального образования

САМАРСКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ

Д.Н. Цивинский

ПРИМЕНЕНИЕ МЕТОДА ЦЕНТРАЛЬНОГО КОМПОЗИЦИОННОГО ПЛАНИРОВАНИЯ В НЕФТЕГАЗОВОМ ДЕЛЕ

Допущено учебно-методическим объединением вузов Российской Федерации по нефтегазовому образованию в качестве учебного пособия для подготовки дипломированных специалистов по направлению 650700 «Нефтегазовое дело» и бакалавров и магистров направления 553600 «Нефтегазовое дело»

Самара 2002

УДК 519.22:681.3(076.5)

Применение метода центрального композиционного планирования в нефтегазовом деле: Учеб. пособ. /*Д.Н. Цивинский*. Самар. гос. техн. ун-т. Самара, 2002, 92 с.

Рассматривается применение метода центрального композиционного планирования в нефтегазовом деле к решению технологических задач строительства скважин. К ним относятся задачи поиска оптимальных условий проводки скважины, получения математической модели процесса без исследования механизма или описания вида "свойство=f(состава)" и многие другие. Описаны свойства оценок неизвестных параметров уравнения, теоретические основы получения квадратного уравнения множественной регрессии, построения матрицы плана, вычисления коэфоициентов уравнения и проведения регрессионного анализа. В достаточном объёме рассмотрен вопрос проверки гипотез. Для облегчения понимания подробно рассмотрены более 50 специальных терминов, причём определение каждого термина представлено отдельной статьёй. Примеры расчёта иллюстрируют все этапы работы по ортогональному центральному композиционному плану и ротатабельному плану второго порядка Бокса-Хантера.

Учебное пособие предназначено для самостоятельной работы студентов, а также может быть полезно аспирантам и инженерам при практической обработке промысловых данных.

ISBN 5-7964-0245-5.

Ил. 8. Табл. 12. Библиогр. 15 назв.

Печатается по решению редакционно-издательского совета
Самарского государственного технического университета

Рецензенты: В.А.Акулов.
В.К.Давыдов

ISBN 5-7964-0245-5

Условные обозначения

Условное обозначение	Физическая величина
$\boldsymbol{B}$	Формальный параметр или вектор параметров
b_j	Оценка j-того коэффициента уравнения регрессии (оценка генерального параметра β_j)
cov_{xy}	Ковариация величин x и y
D	Детерминант матрицы
$\mathrm{D}X$	Математическое ожидание дисперсии случайной величины X
F	Критерий Фишера
$F(x)$	Функция распределения вероятностей случайной величины
i	Номер наблюдения (эксперимента), номер элемента массива, номер строки матрицы наблюдений
j	Номер коэффициента уравнения регрессии, номер фактора, номер столбца матрицы наблюдений
k	Количество факторов (размерность факторного пространства)
l	Число связей, накладываемых на выборку (количество коэффициентов в уравнении регрессии, количество параметров, определённых по выборке)
m_x, m_y	Оценки математических ожиданий случайных величин X, Y
n	Размерность массива данных, число наблюдений (опытов), число *точек*
P	Вероятность события
P	Доверительная вероятность
$p(x)$	Плотность распределения вероятностей случайной величины
p_i	Вероятность i-того события
S^2	Сумма квадратов отклонений
s_x, s_y	Выборочные квадратичные отклонения переменных X, Y (стандартные отклонения)
$s^2_{\text{ад}}$	Дисперсия адекватности
$s^2_{\text{оп}}$	Дисперсия воспроизводимости
s^2_x, s^2_y	Выборочные дисперсии (несмещённые оценки генеральных параметров, σ^2_x, σ^2_y)

Условное обозначение	Физическая величина
t	Критерий Стъюдента
X, Y	Случайные величины
$\mathbf{X}$	Матрица независимых переменных x_{ij}
$\bar{x}$, $\bar{y}$	Выборочные средние значения переменных (оценки математических ожиданий M_x, M_y)
x	Независимая переменная величина (фактор)
$\mathbf{Y}$	Матрица-столбец функции отклика y_i
y	Функция отклика
$\hat{y}$	Значение функции отклика, рассчитанное по уравнению регрессии
$\mathbf{Z}$	Матрица плана экспериментов
z	Кодированное значение переменной величины
α	Уровень значимости, величина звёздного плеча
B	Вектор неизвестных параметров
β_j	Математическое ожидание j-того коэффициента уравнения регрессии (генеральный j-тый параметр)
Γ	Гамма функция
Δ	Интервал варьирования переменных
EX	Математическое ожидание случайной величины X
ε	Малое число
Θ	Статистика критерия
M_x, M_y	Математические ожидания случайных величин X, Y
ν	Число степеней свободы
ρ	Радиус сферы в многомерном пространстве
$б^2{}_x$, $б^2{}_y$	Генеральные дисперсии случайных величин X, Y
$б_x$, $б_y$	Генеральные квадратичные отклонения
υ	Характеристика процесса (неизвестная функция)
Φ	Сумма квадратов разностей экспериментальных и расчётных значений функции отклика

ВВЕДЕНИЕ

В науке и технологии достаточно часто возникает задача поиска оптимальных условий проведения процесса или получения математической модели процесса без исследования механизма. Другими словами, необходимо решить задачу пространственно-временной организации системы при неполном знании механизма явлений, происходящих в системе. Например, при оптимизации химико-технологического процесса найти такое сочетание температуры, давления, концентраций реагентов, структуры потоков и др., чтобы получить максимальный выход целевого продукта. Другой пример: физическая сущность взаимодействия компонентов буровых и цементных растворов, используемых при строительстве скважин, достаточно сложна. Для целей практического приготовления этих растворов в полевых условиях достаточно уравнений, связывающих важные характеристики (пластическая вязкость, напряжение сдвига, водоотдача и др.) с составом.

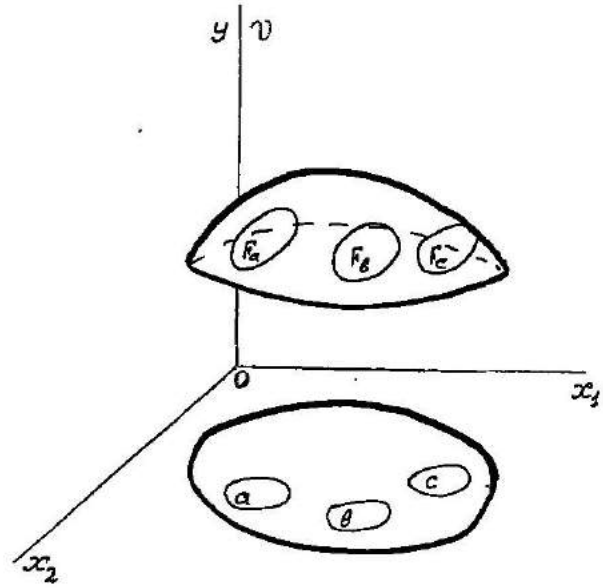
Эти и множество других подобных задач можно решить по разному. Можно провести исследование физической сущности процессов, происходящих в системе, выявить связи и количественные зависимости, создать теорию и математическое описание процесса. Это путь научного познания природы и процессов, происходящих в ней, он достаточно длителен и дорог. Экспериментатор выбирает тот или иной путь исследования, основываясь на своём опыте и интуиции. Но, как правило, системы, подлежащие автоматическому регулированию или оптимизации, оказываются столь сложными, что не поддаются экспериментальному и теоретическому изучению в разумные сроки.

В большинстве случаев задачи автоматизации системы и оптимизации процесса могут быть решены экспериментальным путём при ограниченном знании механизма явлений в системе. В первом случае формальная математическая модель системы необходима, во втором - без неё можно обойтись с помощью *экстремальных экспериментов*.

Математическая теория экстремальных экспериментов стала развиваться с начала пятидесятых годов. Она позволяет выбрать оптимальную стратегию исследования при неполном знании механизма процесса и относительно быстро найти оптимальные условия процесса. Оптимальные условия проведения процесса можно найти и расчётным путём, для чего надо получить уравнение поверхности и с его помощью найти условия экстремума. Теория планирования эксперимента позволяет получить математическую модель процесса или системы, пригодную для практических целей, например поиск оптимальной композиции системы, поиск оптимальных условий процесса, автоматизация процесса и др., с минимальными затратами времени и средств. К достоинствам этого пути можно отнести сравнительно небольшой объём опытов и относительную простоту обработки данных, а к недостаткам - крайне ограниченный физический смысл параметров полученного уравнения, применимость только в изученной области изменения независимых переменных и незначительную роль самого уравнения в развитии научной теории исследованного процесса. Можно также упомянуть тот факт, что при получении неадекватного уравнения большая часть экспериментов объявляется неудачной. Но это другой, отдельный разговор.

Другими словами, пользуясь методами планирования эксперимента, можно исследовать гиперповерхность неизвестной функции, "перемещаясь" по ней в процессе исследования, и экспериментально определить экстремум. При этом на каждом этапе исследования необходимо выбрать оптимальное, в некотором смысле, сочетание условий эксперимента для того, чтобы получить некоторое представление о форме поверхности функции. Можно спланировать эксперименты в интересующей области, получить уравнение поверхности и полученное уравнение использовать для решения задач АСУ и оптимизации.

Задача получения уравнения поверхности кажется простой только на первый взгляд. Рассмотрим некоторую неизвестную зависимость $\upsilon=f(x_1, x_2)$, представленную на рисунке (возможные экспериментальные значения функции отклика y на рисунке не показаны). Очевидно, что поверхность, соответствующая функции υ во всей выделенной области x_1 и x_2, представляет собой купол, весьма далека от плоскости и, соответственно, не может быть аппроксимирована линейным уравнением. Части купола F_a, F_b и F_c вследствие неизбежных ошибок при экспериментальном определении величин y могут быть аппроксимированы линейными уравнениеми, но их вид и/или параметры в общем случае будут разными. Начиная исследование экспериментатор, в общем случае, не имеет представления о форме поверхности функции υ и о том, какое место его опыты занимают в пространстве, образованном независимыми переменными и функцией υ. Экспериментатор не знает также того направления факторного пространства, которое в дальнейшем будет представлять преимущественный интерес. Теоретически результат исследования не должен зависеть от того, в каком локальном участке (a, b, c и т.д.) поставлены опыты или все опыты равномерно распределены по всей области возможных значений переменых x_1 и x_2. На практике, не только параметры искомого уравнения могут сильно различаться для участков a, b и c, но и вид уравнения $y=f(x_1, x_2)$ для каждого участка и всей области может быть различным. Но и это ещё не всё. Конечный результат исследования - вид уравнения $y=f(x_1, x_2)$ и его параметры в различной степени зависят от масштаба измерения переменных и направления их варьирования. Другими словами, результат моделирования зависит от места расположения начала координат, масштаба и от вращения системы координат в многофакторном пространстве. Эти (и другие) трудности явились стимулом развития специальных методов планирования экспериментов. Об одном из них - центральном композиционном планировании - и пойдёт речь в настоящем учебном пособии.



Гипотетическая зависимость $\upsilon=f(x_1, x_2)$
в системе координат (x_1, x_2, υ)

1. ПОСТАНОВКА ЗАДАЧИ

Задача получения формального математического описания исследуемого процесса выглядит следующим образом. Нужно получить некоторое выражение для искомой функции

$$y=f(x_1,\ x_2, \ldots,\ x_k), \tag{1.1}$$

где y - *функция отклика*; x_j - независимые переменные или *факторы*; k - количество факторов, влияющих на процесс. Причём собственно вид функции f, как правило, неизвестен. Пространство, образованное независимыми переменными, называется *факторным пространством*, а геометрический образ, соответствующий функции отклика, - *поверхностью отклика*. Термин "функция отклика" принят для того, чтобы отличать функции, имеющие аналитическое выражение и физический смысл, от так называемых экспериментально-статистических зависимостей. Особенность задачи заключается в том, что исследование поверхности отклика производится при неполном знании механизма изучаемых явлений. Естественно считать, что в таком случае аналитическое выражение функции отклика неизвестно, и её представляют полиномом

$$\upsilon = \beta_0 + \sum_{j=1}^{k}\beta_j x_j + \sum_{\substack{l=1\\ m=2\\ l\neq m}}^{k}\beta_{l.m} x_l x_m + \sum_{j=1}^{k}\beta_{jj} x_j^2 \ldots, \tag{1.2}$$

где υ - характеристика процесса, подлежащая оптимизации или регулированию; β_0, β_j, $\beta_{l.m}$, $\beta_{jj}, \ldots$ - коэффициенты уравнения регрессии. Разложение неизвестной функции в степенной ряд эквивалентно представлению её рядом Тейлора:

$$\beta_1 = \frac{\partial f}{\partial x_1}; \quad \beta_2 = \frac{\partial f}{\partial x_2}; \ldots, \quad \beta_k = \frac{\partial f}{\partial x_k};$$

$$\beta_{1.2} = \frac{\partial^2 f}{\partial x_1 \partial x_2}; \quad \beta_{2.3} = \frac{\partial^2 f}{\partial x_2 \partial x_3}; \ldots; \tag{1.3}$$

$$\beta_{1.1} = \frac{1}{2}\cdot\frac{\partial^2 f}{\partial x^2}; \quad \beta_{2.2} = \frac{1}{2}\cdot\frac{\partial^2 f}{\partial x^2}; \ldots.$$

Очевидно, что для определения коэффициентов уравнения регрессии необходимо поставить некоторое количество опытов, т.е. для нескольких комбинаций значений независимых переменных определить значения характеристики процесса υ, которые при этом обозначают латинской буквой y, и подвергнуть эти данные некоторой математической обработке. При этом возникает следующая трудность. В ходе эксперимента неизбежны флуктуации независимых переменных x_1, $x_2, \ldots, x_k$ и неизбежны ошибки при измерении характеристики процесса y. Естественно, что и те и другие ошибки вносят более или менее существенное искажение в наблюдаемую картину, и в процессе обработки данных эти ошибки каким-либо образом необходимо учесть. Следующая трудность заключается в том, что экспериментатор может осуществить только некоторое конкретное, ограниченное число опытов, т.е. осуществить *выборку* из *совокупности* (под совокупностью подразумеваются все возможные опыты в исследуемом факторном пространстве). При этом экспериментатор не всегда имеет представление о форме поверхности отклика и о том, какое место занимают его опыты в факторном пространстве (см. рисунок во введении), поэтому по результатам выборки можно определить только *выборочные* коэффициенты уравнения регрессии, обозначаемые латинскими буквами b_0, b_j, $b_{l.m}$, b_{jj}:

$$\hat{y} = b_0 + \sum_{j=1}^{k} b_j x_j + \sum_{\substack{l=1 \\ m=2 \\ l \neq m}}^{k} b_{l.m} x_l x_m + \sum_{j=1}^{k} b_{jj} x_j^2 \ldots, \qquad (1.4)$$

где $\hat{y}$ - расчётное значение функции отклика. Выборочные коэффициенты b_0, b_j, $b_{l.m}$, b_{jj} ещё называют ***оценками*** коэффициентов β_0, β_j, $\beta_{l.m}$, β_{jj}. Для определения *генеральных коэффициентов* β_0, β_j, $\beta_{l.m}$, β_{jj} необходимо поставить опыты во всех точках факторного пространства (т.е. осуществить *совокупность*), что реально невозможно. Естественно, что соответствие выборочных коэффициентов генеральным будет при удачном распределении условий опытов по объёму факторного пространства и при относительно небольших отклонениях экспериментальных значений y от неизвестной поверхности отклика υ.

Прежде чем переходить к подробному рассмотрению метода центрального композиционного планирования рассмотрим определения некоторых терминов. При первом чтении их можно пропустить.

1.1. Определения некоторых терминов

Адекватность (*франц.* adequat - адекватный < *лат.* adaequatus - соответственный, тождественный, приравненный, равный; *лат.* adaequo - сравнивать, уравнивать) - соответствие, соразмерность, верность, точность, полное соответствие исследуемому предмету. В теории познания термин "адекватность" служит для обозначения верного воспроизведения объективных связей и отношений действительности в представлениях, понятиях и суждениях. В этом смысле *истина* определяется как адекватность мышления бытию.

В моделировании адекватность - количественная характеристика соответствия модели оригиналу. Критерием адекватности является однородность *дисперсий* - дисперсии *воспроизводимости* $s^2_{оп}$ и дисперсии адекватности $s^2_{ад}$:

$$s^2_{оп}=\sum_{i=1}^{n_{оп}}(y_1-\bar{y})^2/(n_{оп}-1); \qquad (1.5)$$

$$s^2_{ад}=\sum_{i=1}^{n}(y_1-\hat{y}_1)^2/(n-l), \qquad (1.6)$$

где $n_{оп}$ - число опытов в *выборке* на *воспроизводимость*; n - число опытов в выборке, осуществляемой с целью получения уравнения *функции отклика*; l - *число связей*, наложенных на выборку (число параметров уравнения, определённых по выборке). Если дисперсии однородны, математическая модель адекватна, если неоднородны, - неадекватна. Более или менее объективным критерием однородности дисперсий является значение опытного критерия Фишера:

$$F^{оп}_{\nu_1,\nu_2}=\frac{S^2_{ад}}{S^2_{оп}}, \qquad (1.7)$$

где ν_1 - *число степеней свободы* числителя; ν_2 - число степеней свободы знаменателя. Поскольку критерий Фишера табулирован, начиная со значения 1, в числителе должна быть большая дисперсия, в знаменателе - меньшая. Практически дисперсия адекватности обычно больше дисперсии воспроизводимости. Это объясняется тем, что дисперсия адекватности включает в себя два вида ошибок: ошибки расчётные, обусловленные приближённым характером математической модели, и ошибки экспериментальные, а дисперсия воспроизводимости (опытная дисперсия) характеризует только ошибки эксперимента. Опытный критерий Фишера сравнивают с табличным значением критерия Фишера F^{α}, взятого с требуемым *уровнем значимости* α: если опытный критерий Фишера меньше табличного, то принимается *нулевая гипотеза* об отсутствии различия между дисперсиями адекватности и воспроизводимости, т.е. $s^2_{ад}$ и $s^2_{оп}$ однородны и, соответственно, модель адекватна эксперименту, если больше - нулевая гипотеза отклоняется, т.е. дисперсии неоднородны, модель неадекватна.

Величина параметрическая - общее название физических *величин*, приведённых к безразмерному виду путём того или иного соотношения с *параметром*. Величина параметрическая получается в результате *нормирования переменных* и *приведения переменных*. Если параметрическая величина образует *систему координат*, то последнюю называют *параметрической* системой координат. См. также *Кодирование переменных*, *Нормализация*, *Стандартизация случайной величины*.

Воспроизводимость в теории и практике экспериментальных исследований - характеристика точности лабораторного или промышленного эксперимента, а также подтверждение результатов тех или иных *наблюдений* в природе и обществе другими исследователями, авторами в другое время, в тех или иных условиях. Обычно воспроизводимость характеризуется количественно (т.е. числом, в процентах или долях единицы), но может характеризовать явление или процесс и качественно. См. также *Адекватность*, *Дисперсия воспроизводимости*.

Выборка - понятие математической статистики, объединяющее результаты каких-либо однородных наблюдений [8,16]. Выборкой в широком смысле слова называется конечная совокупность результатов наблюдений X_1, $X_2, \ldots, X_n$, представляющих собой независимые одинаково распределённые случайные величины. Определённая таким образом выборка называется *случайной*, а её конкретные значения в каждом отдельном случае x_1, $x_2, \ldots,$ x_n - простой выборкой. В узком смысле понятие выборки связано с теорией статистического *выборочного метода* и предполагает наличие некоторой конечной *совокупности*, из которой эта выборка извлекается (рассматриваются, например, повторные и бесповторные выборки). С точки зрения исследователя, осуществляющего экспериментальные исследования с целью моделирования процесса, выборкой будет называться конкретное количество анализов, опытов, измерений и т.п., а под совокупностью будет подразумеваться абстрактная бесконечность возможных анализов, опытов, измерений и т.п.

Дисперсия (< *лат.* disperse - рассеянно, разбросанно, там и сям; dispersio - рассеяние, разбросанность) в математической статистике и теории вероятностей - одна из характеристик распределения вероятностей случайной величины, наиболее употребительная мера рассеяния её значений, т.е. отклонения её от среднего. Дисперсия - *центральный момент* второго порядка. В теории вероятностей дисперсия DX случайной величины X определяется как математическое ожидание $E(X-M_x)^2$ квадрата отклонения X от её математического ожидания $M_x=EX$. Для случайной величины с дискретным распределением дисперсия определяется формулой

$$DX = \sum_{i=1}^{\infty}(x_i-M_x)^2 p_i, \qquad (1.8)$$

где вероятность $p_i=P(X=x_i)$ при условии, что ряд сходится; для случайной величины X с непрерывным распределением, имеющим плотность вероятности $p(x)$, - формулой

$$DX = \int_{-\infty}^{\infty}(x-M_x)^2 p(x)\,dx, \qquad (1.9)$$

если этот интеграл сходится. Дисперсия имеет важное значение в характеристике качества статистической оценки случайной величины. Наряду с дисперсией в качестве меры рассеяния (той же размерности, что и сама случайная величина) используется квадратный корень из дисперсии $б=\sqrt{DX}$, называемый *квадратичным отклонением* X. Если $DX=0$, то случайная величина X принимает с вероятностью 1 единственное значение M_x. Дисперсия обладает свойством минимальности в том смысле, что

$$DX = \min_{-\infty<a<\infty} E(X-a)^2; \qquad (1.10)$$

при этом min достигается при $a=EX$. См. также *Адекватность, Дисперсия адекватности, Дисперсия воспроизводимости.*

Дисперсия адекватности (*от лат.* disperse - рассеянно, разбросанно, там и сям; dispersio - рассеяние, разбросанность и adaequo - сравнивать, уравнивать, adaequatus - приравненный, равный) - количественная характеристика расхождения экспериментальных значений функции отклика и значений функции отклика, рассчитанных по уравнению, параметры которого определены по *выборке*. Подробнее см. *Адекватность.*

Дисперсия воспроизводимости - количественная характеристика точности эксперимента или *воспроизводимости* результатов наблюдений в природе; вычисляется по формулам

$$s^2_{оп} = \frac{1}{n_{оп}-1} \sum_{i=1}^{n_{оп}} (y_i-\bar{y})^2; \tag{1.11}$$

$$\bar{y} = \frac{1}{n_{оп}} \sum_{i=1}^{n_{оп}} y_i, \tag{1.12}$$

где $n_{оп}$ - число опытов на воспроизводимость; $\nu_{оп}=n_{оп}-1$ - *число степеней свободы* дисперсии воспроизводимости. С точки зрения *метода моментов* дисперсия воспроизводимости является центральным моментом второго порядка и может быть вычислена по формулам:
для дискретной случайной величины

$$\mu_2 = \sum_{i=1}^{n} (x_i-M_x)^2 p_i = s_x^2; \tag{1.13}$$

для непрерывной -

$$\mu_2 = \int_{-\infty}^{+\infty} (x_i-M_x)^2 p(x)dx = s_x^2. \tag{1.14}$$

Для практического расчёта дисперсии воспроизводимости более удобна и точна формула

$$s^2_{оп} = \frac{1}{n_{оп}(n_{оп}-1)} \left\{ n_{оп} \sum_{i=1}^{n_{оп}} y_i^2 - \left(\sum_{i=1}^{n_{оп}} y_i \right)^2 \right\}. \tag{1.15}$$

Для оценки достоверности результатов наблюдений одной дисперсии воспроизводимости недостаточно. Корень квадратный из дисперсии воспроизводимости называется *квадратичным отклонением* или *стандартным отклонением*. Начинающие исследователи обычно с трудом развивают интуитивное восприятие численного значения дисперсии или стандартного отклонения. Является ли дисперсия воспроизводимости, равная, например, 77, большой или малой? Что значит *стандартное отклонение* $0,51 \cdot 10^{-4}$? Оказывается, для интерпретации как дисперсии воспроизводимости, так и стандартного отклонения главное не получить численные значения последних, а правильно сравнить дисперсию воспроизводимости с какой-либо другой дисперсией, например *дисперсией адекватности*, или стандартное отклонение умножить на правильно выбранный *критерий Стьюдента*, чтобы получить *доверительный интервал* для неизвестного *математического ожидания.* На начальных этапах исследований стандартное отклонение $\pm s_y$ сравнивают со средним значением выборки, y_{cp}. Если $|s_y| \ll y_{cp}$, то говорят о *значимом* отличии результатов наблюдений от нуля и **предварительную** *оценку* точности эксперимента осуществляют по отношению s_y/y_{cp}. Если это отношение лежит в пределах 3÷4%, то в первом приближении результаты наблюдений считают воспроизводимыми; в противном случае необходимы дополнительные изыскания.

При использовании центрального композиционного планирования дисперсия воспроизводимости обычно определяется по результатам опытов в центре плана:

$$s^2_{оп} = \frac{1}{(n_0-1)} \sum_{i=1}^{n_0} (y_{0,i} - \bar{y}_0)^2; \qquad (1.16)$$

$$\bar{y}_0 = \frac{1}{n_0} \cdot \sum_{i=1}^{n_0} y_{0,i}, \qquad (1.17)$$

где $y_{0,i}$ – результаты опытов в центре плана, а n_0 – количество опытов в центре плана. В случае ортогонального ЦКП у исследователя есть относительная свобода выбора количества опытов в центре плана (величина звёздного плеча α и n_0 взаимосвязаны, см. табл. 2.1). В случае ротатабельного ЦКП у исследователя нет свободы выбора количества опытов в центре плана – величина звёздного плеча α и n_0 определяются критерием оптимальности плана и связаны с числом факторов k, см. табл. 2.5.

Доверительная вероятность - вероятность достоверности принимаемой гипотезы; характеристика надёжности полученной по *выборке оценки* того или иного параметра:

$$P = P\{|\beta - b| \leqslant \varepsilon_p\}, \qquad (1.18)$$

где β - генеральный параметр; b – его оценка; $\varepsilon_p = f(p)$ – ошибка определения генерального параметра; P - вероятность настолько большая, что событие $|\beta - b| \leqslant \varepsilon_p$ можно считать практически достоверным. Очевидно, что диапазон возможных с вероятностью P значений ошибки от замены β на b равен $\pm\varepsilon_p$. Вероятность появления ошибок, больших по абсолютной величине, чем ε_p, или вероятность событий $|\beta - b| \geqslant \varepsilon_\beta$ называется *уровнем значимости:*

$$\alpha = 1 - P = P\{|\beta - b| \geqslant \varepsilon_p\}. \qquad (1.19)$$

Иначе выражение (1.18) может быть интерпретировано как вероятность того, что истинное значение параметра β находится в пределах

$$b-\varepsilon_p<\beta<b+\varepsilon_p, \tag{1.20}$$

где выборочный параметр b - по существу *случайная величина*, а ошибка его определения $(\beta-b)$ в выражениях (1.18) и (1.19) - также случайная величина. Интервал $I_p=b\pm\varepsilon_p$ называется *доверительным интервалом*. Границы интервала $b_{min}=b-\varepsilon_p$ и $b_{max}=b+\varepsilon_p$ называется *доверительными границами*. Доверительный интервал при принятой доверительной вероятности определяет точность оценки. Величина доверительного интервала зависит от принимаемой доверительной вероятности, т.е. от той вероятности, с которой гарантируется нахождение искомого параметра β внутри доверительного интервала, другими словами, чем выше гарантия надёжности оценки, тем больше величина интервала, в котором может находиться генеральный параметр. См. также *Уровень значимости*.

Доверительное отклонение - *функция* от результатов наблюдений и *доверительной вероятности*, позволяющая оценить *доверительный интервал*, который с вероятностью $P=1-\alpha$ "накрывает" неизвестное значение параметра:

$$\frac{s_x}{\sqrt{n}}\cdot t_\nu^\alpha, \tag{1.21}$$

где s_x - *квадратичное (стандартное) отклонение*; n - количество наблюдений в выборке; t - случайная величина, зависящая только от *числа степеней свободы* ν выборочной дисперсии и *уровня значимости* α, называемая *критерием Стьюдента*. Необходимо отметить, что доверительное отклонение не зависит ни от *математического ожидания* M_x, ни от *генерального параметра* $б_x$. Это *случайная величина*, зависящая только от квадратичного отклонения s_x и принимаемого исследователем *уровня значимости* α. См. также *Доверительная вероятность*, *Доверительные границы*.

Доверительные границы - см. *Доверительная вероятность*, *Доверительное отклонение*, *Доверительный интервал*.

Доверительный интервал - статистическая оценка параметра исследуемого вероятностного распределения, имеющая вид интервала, границами которого служат функции от результатов наблюдений и *доверительной вероятности* и который с вероятностью P "накрывает" неизвестное значение параметра. Дело в том, что значение оценки в каждом конкретном случае может отличаться от *истинного* значения параметра (*математического ожидания*), и, следовательно, в интерпретации результатов эксперимента имеется некоторая доля неопределённости. При грубых оценках величина этой неопределённости выражается с помощью *выборочной дисперсии* или *квадратичного отклонения*, т.е. вполне возможно, что неизвестный *генеральный параметр* χ находится в интервале $x_{cp}\pm s_x$; также возможно, что он находится в интервале $x_{cp}\pm 2s_x$ и т.д. Другими словами, для верной интерпретации результатов эксперимента важно установить для оценки генерального параметра χ **интервал** вместо отдельной точки x_{cp}, причём хотя бы одна точка этого интервала, а именно x_{cp}, и рассматривалась бы как "наилучшая" оценка для χ. Задача определения доверительного интервала решалась бы достаточно просто, если бы был известен закон распределения оценки x_{cp} (1.11):

$$P\{|M_x-\bar{x}|\leqslant\varepsilon_p\} = \int_{-\varepsilon_p}^{+\varepsilon_p} p(x)\,dx = P. \tag{1.22}$$

Также просто решалась бы эта задача при известной генеральной дисперсии σ^2_x (знание генеральной дисперсии σ^2_x позволяет оценивать доверительный интервал даже по одному наблюдению), но, к сожалению, генеральную дисперсию σ^2_x невозможно получить из наблюдений, её можно только оценить при помощи выборочной дисперсии s^2_x. Например, для выборки объёма n

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i; \qquad s^2_x = \frac{1}{n-1}\cdot\sum_{i=1}^{n}(x_i-\bar{x})^2,$$

где x_{cp} - оценка математического ожидания M_x, а s^2_x - оценка генеральной дисперсии σ^2. Известно, что нормальное распределение наиболее широко распространено в природе. Задача оценки доверительного интервала для математического ожидания нормально распределённой случайной величины была решена в 1908 г. У. Госсетом, известным под псевдонимом Student:

$$\bar{x} - \frac{s_x}{\sqrt{n}}\cdot t^{\alpha}_{\nu} \leqslant m_x \leqslant \bar{x} + \frac{s_x}{\sqrt{n}}\cdot t^{\alpha}_{\nu}, \tag{1.23}$$

где t - критерий Стьюдента, случайная величина, зависящая только от *числа степеней свободы* ν выборочной дисперсии и *уровня значимости* α. Из оценки видно, что уменьшение доверительного интервала обратно пропорционально корню квадратному из числа наблюдений; другими словами, для того, чтобы повысить точность результатов наблюдений в два раза, необходимо увеличить объём выборки в четыре раза. См. также *Доверительная вероятность, Доверительное отклонение, Доверительные границы, Стьюдента критерий.*

Значимость коэффициента - см. *Доверительное отклонение, Квадратичное отклонение, Стандартное отклонение, Стьюдента критерий.*

Инвариант (*фр.* invariant - инвариантный, неизменный, инвариант, неизменная величина; от лат. invarians, род. падеж invariantis - неизменяющийся) - (*мат.*) "отображение φ рассматриваемой совокупности m математических объектов, снабжённой фиксированным отношением эквивалентности ρ, в другую совокупность n математических объектов, постоянное на классах эквивалентности m по ρ". Проще, но нестрого - выражение, остающееся неизменным при определённом преобразовании переменных, связанных с этим выражением, например преобразование системы координат, вращение системы координат. См. также *Кодирование переменных, Ротатабельность, Униформ-ротатабельное планирование.*

Квадратичное отклонение, *квадратичное уклонение*, величин x_1, $x_2, \ldots, x_n$ от a - квадратный корень из выражения

$$\frac{(x_1-a)^2 + (x_2-a)^2 + \ldots + (x_n-a)^2}{n}. \tag{1.24}$$

Наименьшее значение квадратичное отклонение имеет при $a=x_{cp}$, где x_{cp} - среднее арифметическое величин $x_1, x_2, \ldots, x_n$:

$$x_{cp} = \bar{x} = \frac{x_1 + x_2 + \ldots + x_n}{n}. \tag{1.25}$$

В теории вероятностей *квадратичное отклонение* $б_X$ случайной величины X (от её математического ожидания) определяют как квадратный корень из дисперсии $\sqrt{DX}$ и называют также *стандартным отклонением* величины X. Для любой случайной величины X с математическим ожиданием M_X и квадратичным отклонением $б_X$ вероятность отклонений X от M_X, больших по абсолютной величине $kб_X$, $k>0$, не превосходит $1/k^2$. В случае нормального распределения указанная вероятность при $k=1$ равна 0,3174, при k=2 - равна 0,0456 и при k=3 - равна 0,0027. В практических задачах, приводящих к нормальному распределению, отклонения больше, чем утроенный стандарт (*квадратичное отклонение*), практически невозможны, или, другими словами, на практике пренебрегают возможностью отклонений от среднего, больших $3\sigma_x$ (правило трёх сигма).

В математической статистике квадратичное отклонение употребляют как меру качества статистических оценок и называют в этом случае *квадратичной погрешностью* (ошибкой). См. также *Стандартное отклонение*.

Ковариация - количественная мера той неопределённости, которая возникает в силу того, что коэффициенты уравнения регрессии определяются не независимо друг от друга. Коэффициенты ковариаций расположены на пересечениях i-тых строк и j-тых столбцов корреляционной матрицы $(X^TX)^{-1}$. Их величина характеризует степень корреляции коэффициентов b_j в пассивном эксперименте [18], при ортогональном планировании корреляционная матрица $(X^TX)^{-1}$ диагональна - кроме главной диагонали все члены равны нулю (например, в полном факторном эксперименте [17]), при *униформ-ротатабельном планировании* ковариация коэффициентов уравнения регрессии определяется выполнением условия постоянства информационного потенциала плана в возможно большем объёме факторного пространства (*далее по тексту раздела* 2.2). См. также *Корреляция*.

Кодирование переменных в задачах *планирования эксперимента* - переход от физических переменных к безразмерным *кодированным*, причём независимо от физической сущности минимальным значениям всех физических переменных соответствует значение -1, а максимальным значениям - значение +1.

Идея кодирования переменных принадлежит английскому статистику и генетику Р. Фишеру который в начале 20-х годов обратил внимание на то, что диагональность матрицы $\boldsymbol{X}^T\boldsymbol{X}$ позволит получить некоррелированные коэффициенты b_j, а необходимым и достаточным условием диагональности матрицы $\boldsymbol{X}^T\boldsymbol{X}$ является ортогональность матрицы $\boldsymbol{X}$. Другими словами, Р. Фишер предложил спланировать эксперимент таким образом, чтобы исходная матрица $\boldsymbol{X}$ была ортогональна. Развитие идеи планирования эксперимента привело к тому, что Д. Бокс и К. Уилсон предложили все физические переменные представлять только в виде +1 и -1. При таком подходе для каждой физической переменной (фактора) выбирается область допустимых значений и максимальным значениям всех факторов, независимо от физической сущности и величины, присваивается значение +1 (*верхний уровень*), а минимальным значениям, соответственно, -1 (*нижний уровень*).

В общем случае центр плана (*нулевой уровень*) по каждому фактору

$$x_j^0 = \frac{x_j^{max} + x_j^{min}}{2}; \tag{1.26}$$

а интервал варьирования:

$$\Delta x_j = \frac{x_j^{max} - x_j^{min}}{2}; \tag{1.27}$$

Основная идея кодирования переменных по Д. Боксу и К. Уилсону заключается в том, что независимо от величин физических переменных x_1, $x_2, \ldots,$ x_k и, соответственно, Δx_j при переходе в кодированную систему координат z_1, $z_2, \ldots,$ z_k интервал варьирования Δz_j для всех факторов приравнивается единице:

$$\Delta x_j \Rightarrow \Delta z_j = 1; \; j = 1, \; 2, \ldots, \; k; \tag{1.28}$$

В факторном пространстве размерности $k+1$ точка с координатами $x^0{}_1$, $x^0{}_2, \ldots,$ $x^0{}_k$ называется центром плана. При переходе к кодированным переменным z_1, $z_2, \ldots,$ z_k в эту точку без поворота переносится начало координат и в этой новой (кодированной) системе координат z_1, $z_2, \ldots,$ z_k все независимые переменные изменяются в интервале от -1 до +1:

$$z_j = \frac{x_j - x_j^0}{\Delta x_j}. \tag{1.29}$$

Другими словами, под кодированием переменных подразумевается переход из физической системы координат в безразмерную систему координат, в которой все переменные изменяются от -1 до +1, а начало координат расположено в центре правильного многогранника с 2^k вершинами, где k - количество независимых переменных. Исключение составляет только функция отклика y, которая измеряется и участвует в последующих расчётах в прежних физических единицах. Не следует смущаться отсутствием строгости в рассуждениях "факторное пространство кодированных переменных" и то же пространство с физическими переменными y_1, y_2, ..., y_n, – в данном случае это не более, как методический приём. Не следует путать с *Нормированием переменных*.

Корреляция (от позднелат. correlatio – соотношение) – вероятностная или статистическая зависимость между двумя и более случайными величинами. В отличие от функциональной зависимости корреляция возникает тогда, когда зависимость одного признака (случайной величины Y) от другого (от другой случайной величины X) осложняется случайными *факторами* или когда среди условий, от которых зависят две случайные величины, например Y и Z, имеются общие для них обоих условия X_1, X_2, ... Такого рода зависимости иногда выявляются визуально с помощью графиков и последующего *регрессионного анализа*, иногда корреляционная зависимость не очевидна, и её выявляют с помощью корреляционного анализа. Необходимо отметить, что статистическая зависимость, как бы ни была она сильна, никогда не может установить причинной связи: предположения о *причинах* и *следствиях* следует искать вне статистики, например в сфере физических сущностей явления.

В основе математической теории корреляции лежит предположение о том, что все явления в природе в большей или меньшей мере подчинены определённым вероятностным закономерностям. Зависимость между двумя случайными событиями проявляется в том, что условная вероятность наступления одного из них при наступлении другого отличается от безусловной вероятности. Или, другими словами, влияние одной из них (например X) на другую (например Y) не вполне конкретно, не жёстко функционально, хотя при возрастании случайной величины X, другая, Y, имеет тенденцию возрастать или убывать. Это характеризуется тем, что при каждом фиксированном значении X в результате наблюдений получается множество значений Y, характеризующееся некоторым распределением вероятностей. В таких случаях *функция* $y=f(x)$ называется *регрессией* величины Y по X, а график такой зависимости – *линией регрессии* Y по X.

Термин "корреляция" используется также для констатации взаимосвязи между параметрами математической модели, которые вычисляются по экспериментальным данным. В этом случае обычно не идёт речь о поиске той или иной функциональной зависимости между вычисляемыми параметрами, важно наличие или отсутствие корреляции. В случае наличия корреляции между параметрами уравнения, представляет определённый интерес относительная величина корреляции между теми или иными параметрами.

В случае обработки данных методом пассивного эксперимента все коэффициенты уравнения регрессии взаимосвязаны. Степень взаимосвязи между коэффициентами характеризуют недиагональные элементы a_{ij} корреляционной матрицы $(X^{T}X)^{-1}$, где $i \neq j$. Эти элементы называются *ковариациями*. Ковариации являются количественной мерой той неопределённости, которая возникает в силу того, что коэффициенты уравнения регрессии определяются не независимо друг от друга [2. 18]. При ортогональном планировании корреляционная матрица $(X^{T}X)^{-1}$ диагональна - члены главной диагонали характеризуют дисперсии коэффициентов b_j, а все остальные члены равны нулю. В результате все коэффициенты уравнения регрессии определяются независимо друг от друга и все ковариации оказываются равными нулю [2. 10]. Ортогональными планами являются полный факторный эксперимент [2, 10, 17] и ортогональное центральное композиционное планирование (см. разд. 2.1, [2, 10]).

При ротатабельном планировании эксперимента обсуждаются зависимости коэффициентов корреляции *эффектов* парного взаимодействия ($b_{1,u}$ и $b_{1,m}$) и квадратичных эффектов (b_{11} и b_{mm}) от поворота системы координат в факторном пространстве от 0° до 90° (см. рис.2.5, а, б). Выполнение условия постоянства информационного потенциала плана в возможно большем объёме факторного пространства, $0<\rho<1$, приводит к тому, что в отличие от ортогонального планирования в ротатабельных униформ-планах нет свободы выбора величины звёздного плеча α и числа опытов в центре плана n_0. При этом свободный член b_0 коррелирован с квадратичными эффектами b_{jj}, а квадратичные эффекты коррелированы не только со свободным членом, но и между собой.

Критерий (от греч. ϰριτηριο - критерий, признак, по которому можно судить верно) - мерило для опредления достоверности, соответствия человеческого знания объективной реальности, а также признак. на основании которого производится оценка, определение или классификация чего-либо, мерило оценки.

Линейности условие метода НК (*лат.* linearis - состоящий из линий, штриховой; линейный, геометрический) - условие, согласно которому в уравнении регрессии $y=f(x_1, x_2, ..., x_k)$, предполагаемом для описания неизвестной функции $\upsilon=\varphi(x_1, x_2, ..., x_k)$, каждая оценка неизвестного параметра β_j встречается один раз и только в первой степени. Например:

$$y=b_0+b_1x_1+b_2x_2+\ldots+b_kx_k. \qquad (1.30)$$

Подробнее. Модель $y=f(x_1,\ x_2,\ldots,\ x_k)$ должна быть *линейна относительно параметров* b_j, но не обязана быть линейной относительно $x_1,\ x_2,\ldots,\ x_k$. Дело в том, что, по существу, метод НК даёт *несмещённые оценки* параметров, являющихся линейными функциями от наблюдений и имеющих минимальную дисперсию. Другими словами, существенна линейность только относительно параметров, тогда как, возможно, более "естественная" линейность *функции отклика* y от $x_1,\ x_2,\ldots,\ x_k$ с точки зрения метода НК принципиального значения не имеет. Таким образом, модель линейной регрессии включает в себя все виды полиномиальной зависимости y от $x_1,\ x_2,\ldots,\ x_k$. Например, непосредственно полиномиальная зависимость

$$y=b_0+b_1x+b_2x^2+b_3x^3+\ldots+b_nx^n, \qquad (1.31)$$

нелинейна относительно x, но линейна относительно β_j и с точки зрения метода НК является уравнением линейной регрессии. Аналогично в случае нескольких переменных квадратное уравнение

$$y=b_0+b_1x_1+b_2x_2+\ldots+b_kx_k+b_{12}x_1x_2+b_{23}x_2x_3+\ldots+b_{k-1,k}x_{k-1}x_k+$$
$$+b_{11}x_1^2+b_{22}x_2^2+\ldots+b_{kk}x_k^2 \qquad (1.32)$$

с точки зрения *метода* НК представляет собой модель линейной множественной регрессии, несмотря на то, что наличие парных произведений и квадратных членов по существу и обеспечивают удовлетворительное описание криволинейных гиперповерхностей. С другой стороны, уравнение линейное относительно x_j

$$y=b_0+b_1x_1+b_1^2x_2+b_2x_3, \qquad (1.33)$$

с точки зрения *метода* НК линейным не является, поскольку в него входят и b_1 и $b^2{}_1$. Аналогично модель, нелинейная по своей сущности

$$y=b_0+b_1x_1+b_2x_1^2+b_3\sin x_1\cos x_2, \qquad (1.34)$$

с точки зрения *метода* НК является линейной. Причина кажущегося противоречия - "модель линейная относительно b_j, нелинейная относительно x_j" и наоборот - кроется в изначальной постановке задачи оценки неизвестных параметров эмпирического распределения методом НК. Эта задача решается путём дифференцирования некоторой функции

$\Phi=f(b_j, x_1, y_1)$ по оценкам параметров b_j искомого уравнения, причём предполагается, что каждая оценка b_j в функции Φ встречается один раз и только в первой степени.

Причина фактической неопределённости поиска уравнения, описывающего неизвестную зависимость $\upsilon=\varphi(x_1, x_2, \ldots, x_k)$, кроется в том, что каждое событие в мире является результатом одновременного воздействия некоторого множества детерминированных факторов (т.е. факторов, имеющих определённую причинно-следственную связь). Случайные комбинации детерминированных факторов в тот или иной момент времени приводят в целом к случайному характеру события как по силе (интенсивности), так и по времени его осуществления. Именно это и обуславливает невозможность точного определения неизвестных параметров и необходимость какой-либо **оценки** этих параметров. Но задача собственно оценки неизвестных параметров отступает на второй план перед проблемой ***априорного принятия вида функции.*** Диалектика научного познания проявляется в том, что, имея массив наблюдений, исследователь на основе знаний, опыта, интуитивных соображений предполагает вид зависимости функции отклика от факторов и только после этого каким-либо методом вычисляет оценки параметров ***принятой функции.*** Если в результате оказывается, что принятая функция (математическая модель) неадекватна, т.е. не соответствует результатам эксперимента при достигнутой точности опытов, то у исследователя есть три выхода: попробовать другой метод обработки данных, предположить другой вид функциональной зависимости $y=f(x_1, x_2, \ldots, x_k)$ или повторить (продолжить) эксперименты. Например, потерпев неудачу при обработке данных по уравнению $y=b_0+b_1x$, попробовать уравнение вида $y=b_0+b_1/x$, другое $y=x/(b_0+b_1x)$ и т.д.

Вообще известно несколько методов оценки неизвестных параметров, все они различаются по точности и, что немаловажно, по трудоёмкости. Наибольшее распространение получил метод НК, который является частным случаем принципа максимума правдоподобия (МП), общего метода построения оценок неизвестных параметров исследуемого распределения. Несмотря на то, что метод НК, в общем случае, не обладает даже асимптотическими оптимальными свойствами, он обладает своими собственными оптимальными свойствами и в то же время совпадает с методом МП в важном случае нормально распределённых наблюдений (нормальное распределение наиболее широко распространено в природе). См. также *Линеаризация, Наименьших квадратов метод, Преобразование факторного пространства.*

Математическое ожидание, среднее значение, - одна из важнейших числовых характеристик распределения *вероятностей* случайной величины. Для случайной величины X, принимающей последовательность значений x_1, $x_2, \ldots, x_i, \ldots$ с вероятностями, равными соответственно p_1, $p_2, \ldots, p_i, \ldots$, *математическое ожидание* определяется:

для дискретной случайной величины при условии, что ряд сходится абсолютно, формулой

$$M_x = \sum_{i=1}^{\infty} x_i p_i; \tag{1.35}$$

для случайной величины X с непрерывным распределением, имеющим плотность вероятности $p(x)$, если интеграл сходится абсолютно, формулой

$$M_x = \int_{-\infty}^{\infty} x p(x)\, dx. \tag{1.36}$$

Математическое ожидание характеризует наиболее вероятное расположение значений случайной величины.

Название "математическое ожидание" происходит от понятия "ожидаемого значения выигрыша" (математического ожидания выигрыша), впервые появившегося в теории азартных игр в трудах Б. Паскаля и Х. Гюйгенса в XVII в. Но впервые в полной мере это понятие было оценено и использовано П.Л. Чебышевым (сер. XIX в.); термин "Математическое ожидание" ввёл П. Лаплас (1795).

Незначимость коэффициента - см. *Доверительное отклонение*, *Квадратичное отклонение*, *Стандартное отклонение*, *Стьюдента критерий*.

Несмещённая оценка - статистическая оценка параметра распределения вероятностей по результатам наблюдений, лишённая систематической ошибки. Например, если результаты наблюдений x_1, $x_2, \ldots, x_n$ являются взаимно независимыми случайными величинами, имеюшими одинаковое нормальное распределение, заданное плотностью

$$p(x) = \frac{1}{\sqrt{2\pi б^2}} \cdot e^{-(M_x - x)^2 / 2б^2} \tag{1.37}$$

с неизвестными параметрами M_x и $б^2$, то среднее арифметическое:

$$\overline{x} = (x_1 + x_2 + \ldots + x_n)/n \tag{1.38}$$

будет несмещённой оценкой для M_x. Используемая ранее для оценки $б^2$ выборочная дисперсия

$$s_x^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2 \tag{1.39}$$

не является несмещённой оценкой, так как среднее арифметическое само зависит от элементов выборки. Для устранения *смещения оценки* нужно *число степеней свободы* в выражении для $s^2{}_x$ уменьшить на единицу. Несмещённой оценкой для $б^2$ служит

$$s_x^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2. \tag{1.40}$$

См. также *Дисперсия воспроизводимости.*

Нормализация – 1) установление *нормы,* образца; 2) приведение к норме, к нормальному состоянию, урегулирование; 3) вид термической обработки стали; 4) нормализация или *нормирование переменных* - преобразование *переменных* (температура, давление, концентрация и т.д.) к нормированному виду при котором все переменные, независимо от физической *сущности* и *величины,* изменяются в пределах от нуля до единицы или нескольких единиц. См. также *Величина параметрическая, Кодирование переменных, Приведение переменных.*

Нормирование переменных - частный случай преобразования *физических величин* к безразмерному виду путём деления исходной переменной величины на аналогичную постоянную; 1) *локализация* области *варьирования фактора* производится путём нормирования переменных, при этом *минимальное* значение фактора принимается за ноль, а *максимальное* - за единицу. В результате безразмерная нормированная величина изменяется в интервале от нуля до единицы:

$$z_j = \frac{x_j - x_j^{min}}{x_j^{max} - x_j^{min}}, \tag{1.41}$$

Например, в задачах оптимизации технологических процессов нормирование переменных позволяет перейти в *факторное пространство,* в котором все переменные находятся в положительной области и изменяются от нуля до единицы. В задаче получения статистических характеристик структуры потоков для приведения концентрации технологического трассера к безразмерному виду постоянной величиной в знаменателе является его начальная концентрация; 2) Нормирование *случайных величин* производится путём деления их на *квадратичное отклонение*. См. также *Величина параметрическая, Нормализация, Приведение переменных*.

Нулевая гипотеза - гипотеза об отсутствии различия. См. также *Статистических гипотез проверка, Стьюдента критерий*.

Однородность дисперсий выборочных - факт принадлежности соответствующих *выборок* к одной и той же *совокупности*. Однородность дисперсий проверяется с помощью *Фишера критерия*. См. также *Адекватность, Дисперсия адекватности, Статистических гипотез проверка*.

Опытная дисперсия - см. *Дисперсия воспроизводимости*.

Ортогональное центральное композиционное планирование - *метод* получения *уравнения регрессии* второго порядка, в котором за *критерий оптимальности* плана принят критерий *ортогональности* (отсутствие *корреляции* между *коэффициентами* уравнения регрессии). Коэффициенты уравнения регрессии, полученного в результате *ортогонального* центрального композиционного планирования, определяются независимо друг от друга, но с разной степенью точности. Более того, при повороте *координатных* осей *факторного пространства* величины дисперсий коэффициентов меняются по весьма сложному закону. Другими словами, точность уравнения зависит от выбранного исследователем места расположения центра плана, от радиуса ядра плана ρ и от направления координатных осей в факторном пространстве. *Далее по тексту*.

Ортогональность (*греч.* **ορθογωνιος** - прямоугольный) - обобщение понятия перпендикулярности. Если два *вектора* в трёхмерном пространстве перпендикулярны, то их скалярное произведение равно нулю. Понятие перпендикулярности векторов распространяется и на многомерное пространство, в котором определено скалярное произведение, в частности на *пространство факторное* при *планировании эксперимента*. Так, *матрица плана* эксперимента называется ортогональной, если скалярные произведения всех пар вектор-столбцов равны нулю:

$$\sum_{i=1}^{n} z_{i,l} z_{i,m} = 0; \quad l \neq m; \ l,\ m = 0,\ 1,\ 2, \ldots,\ k,$$

где i - номер строки; l, m - номера столбцов; n - число строк; k - число столбцов матрицы наблюдений. Важным следствием ортогональности матрицы плана эксперимента является независимость коэффициентов уравнения регрессии друг от друга.

Термин "ортогональность" встречается у Евклида (III в. до н.э.).

Оценка - количественная характеристика *генерального параметра*, получаемая по результатам *выборки*. К оценкам параметров предъявляется комплекс требований. Важнейшие среди них - *несмещённость, состоятельность* и *эффективность*. См. также *Выборочный метод, Дисперсия, Математическая статистика*.

Параметр (от греч. **παραμετρεω** - мерить что-либо, сопоставляя его с чем-либо; измерять что-либо по чему-либо; сравнивать что-либо по чему-либо) - величина, значения которой служат для различения элементов некоторого множества между собой. В зависимости от конкретного множества различают следующие параметры.

В математической статистике параметр - характеристика совокупности, например математическое ожидание, дисперсия. Параметр совокупностей обычно обозначают греческими буквами, в отличие от их оценок, вычисляемых по результатам выборок и обозначаемых латинскими буквами.

В математическом моделировании параметр - величина, значения которой служат для конкретизации той или иной математической модели, например уравнения Антуана и Андраде математически изоморфны, но первое описывает температурную зависимость давления насыщенных паров жидкости, а второе - коэффициента динамической вязкости:

$$p = \exp\left(A + \frac{B}{T+C}\right); \qquad \mu = \exp\left(A + \frac{B}{T+C}\right),$$

где в первом случае B - аналог теплоты конденсации, а во втором - энергия активации вязкого течения; коэфффициент A - предэкспоненциальный множитель; коэфффициент C - формальный параметр.

Параметр в физическом моделировании и в технике - величина, являющаяся существенной характеристикой системы, технического устройства, явления или процесса. Например, в гидромеханических процессах такими величинами являются динамический коэффициент вязкости жидкой фазы, удельные плотности жидкой и твёрдой фаз, размеры и коэффициент формы частиц твёрдой фазы и др.; для тепловых процессов такими параметрами являются удельные теплоёмкость и теплопроводность, температурный напор и т.д. Параметры могут быть постоянными и переменными (т.е. могут зависеть от времени и/или системы координат).

Планирование (*лат.* planum - равнина, плоскость; planus - плоский, ровный; ясный, понятный) - разработка определённого порядка, последовательности в изложении чего-либо, в исполнении чего-либо. Здесь - вычисление условий проведения экспериментов в *ортогональном центральном композиционном планировании* и в *униформ-ротатабельном планировании.*

Планирование эксперимента - раздел математической статистики, изучающий рациональную организацию измерений, подверженных случайным ошибкам. Обычно рассматривается следующая схема планирования эксперимента. Экспериментатор ограничивает некоторое допустимое множество **X**, в котором переменные x_j могут принимать фиксированные значения, определяемые *критерием оптимальности* плана. При этих фиксированных значениях x_1, $x_2, \ldots, x_k$ со случайными ошибками измеряется функция $f(\mathrm{B}, \boldsymbol{X})$, зависящая от неизвестных параметров В, векторных или числовых, β_0, $\beta_1, \ldots, \beta_k$, и от переменных **X**. Целью эксперимента является обычно либо оценка параметров В или их функций, либо проверка некоторых гипотез о параметрах В. Под *планом эксперимента* понимается совокупность значений, задаваемых переменными x_1, $x_2, \ldots, x_k$ в эксперименте. Критерий оптимальности плана эксперимента формулируется исходя из целей эксперимента. Как правило, оценки параметров В ищут *методом наименьших квадратов*, а гипотезы о параметрах В проверяют с помощью *F*-критерия Фишера. В обоих случаях в качестве критерия оптимальности плана выбирается некоторая функция от дисперсий и коэффициентов корреляции оценок, β_0, $\beta_1, \ldots, \beta_k$, метода наименьших квадратов. В тех случаях, когда $f(\mathrm{B}, \boldsymbol{X})$ линейно зависит от В, оптимальный план часто можно построить до проведения эксперимента, в других случаях уточнение плана эксперимента происходит по ходу эксперимента.

Начало планированию эксперимента положили труды английского биолога Р. Фишера (1935), подчеркнувшего, что рациональное планирование даёт не менее существенный выигрыш в точности оценок, чем оптимальная обработка результатов измерений. Это положило начало разработке методов планирования эксперимента: *полного факторного эксперимента* (плана Бокса первого порядка), композиционных планов Бокса - Уилсона, ротатабельных планов Бокса - Хантера и др. В этих планах функция $f(B, \ \mathbf{X})$ зависит от вектора $\mathbf{X}$ (x_1, x_2, ..., x_k) переменных (факторов) с конечным числом возможных значений и характеризует сравнительный эффект влияния каждого фактора или комбинаций различных факторов на функцию (отклика). Значительное развитие получили методы планирования эксперимента в химии, химической технологии и технике по поиску оптимальных условий протекания того или иного процесса.

Специфическими методами обладает планирование отсеивающих экспериментов, в которых выделяются те компоненты вектора $\mathbf{X}$, которые сильнее всего влияют на функцию $f(B, \ \mathbf{X})$, что очень важно на начальных стадиях исследований, когда вектор $\mathbf{X}$ имеет большую размерность.

Методы планирования экспериментов тесно связаны с теорией приближения функций и математическим программированием. Построены и исследованы свойства оптимальных планов для широкого класса моделей. Разработаны итерационные алгоритмы планирования эксперимента, дающие во многих случаях удовлетворительное решение задач.

Плечо звёздное - величина, определяющая радиус второй *сферы* плана второго порядка, на которой расположены *точки звёздные*. В общем случае величина звёздного плеча определяется выбором *критерия оптимальности* плана. В частном случае *ортогонального центрального композиционного планирования* есть относительная свобода выбора величины звёздного плеча. Эта величина связана с числом опытов в центре плана. В случае *униформ-ротатабельного планирования* свободы выбора величины звёздного плеча нет - эта величина зависит только от числа *факторов*. См. также *Кодирование переменных, Ротатабельность, Ядро плана второго порядка.*

Поверхность отклика - геометрический образ, соответствующий *функции отклика* в *пространстве факторном*. Термин "Поверхность отклика" обычно используется при обсуждении вопросов, связанных с получением экспериментально-статистических моделей, в задачах оптимизации и при планировании экстремальных экспериментов. В таких случаях рассматривается абстрактное (концептуальное) многомерное пространство, образованное k независимыми переменными и функцией отклика. При этом совсем не обязательно, чтобы это $(k+1)$-мерное пространство было образовано путём добавления к трём пространственным координатам времени и других параметров. В качестве факторов могут быть любые независимые переменные x_1, x_2, ..., x_k, главное условие - наличие функциональной связи вида $y=f(x_1, \ x_2, \ldots, \ x_k)$. Особенность задачи заключается в том, что исследование поверхности отклика производится при неполном знании механизма изучаемых явлений. В таких случаях аналитическое выражение функции отклика, как правило, неизвестно и её представляют полиномом (1.2). *Далее по тексту*.

Приведение переменных - общий случай преобразования физических величин к безразмерному виду путём деления исходной переменной величины на аналогичную постоянную. В зависимости от того, какая физическая величина принимается за постоянную, различают *нормирование переменных, кодирование переменных* и приведение к параметрическому виду.

При нормировании переменных производится локализация области варьирования *фактора*, минимальное значение фактора принимается за ноль, а максимальное - за единицу. В результате безразмерная нормированная величина изменяется в интервале только от нуля до единицы.

Нормирование случайных величин производится путём деления их на *квадратичное отклонение*. Получаемое при этом распределение имеет *дисперсию*, равную единице.

При кодировании переменных в задачах *планирования эксперимента* независимо от физической сущности минимальным значениям всех физических переменных присваивается значение -1, а максимальным значениям - значение +1. Такое преобразование факторного пространства позволяет получить систему координат, *инвариантную* к некоторым преобразованиям.

В остальных случаях выбор постоянной, принимаемой за знаменатель, определяется конкретными обстоятельствами, например, в задаче получения статистических характеристик структуры потоков для приведения *времени* к безразмерному виду постоянной величиной в знаменателе является среднее время пребывания жидкости в технологическом аппарате. Приведение температуры к безразмерному виду осуществляется путём деления абсолютной температуры вещества на его критическую температуру. Очевидно, что в отличие от нормированной величины приведённая или *параметрическая величина* может изменяться в достаточно широких пределах. См. также *Нормализация, Стандартизация случайной величины*.

Пространство факторное - абстрактное (концептуальное) многомерное *пространство*, образованное k независимыми переменными и *функцией отклика*. При рассмотрении факторного пространства совсем не обязательно, чтобы это $(k+1)$-мерное пространство было образовано путём добавления к трём пространственным координатам времени и других *параметров*. В качестве *факторов* могут быть любые независимые переменные x_1, $x_2, \ldots,$ x_k, например температура, давление, концентрации реагентов и т.д. Вначале исследователя интересует вид функциональной связи $y=f(x_1,\ x_2, \ldots,\ x_k)$, численные значения параметров и *адекватность* искомого уравнения эксперименту. На следующих этапах решаются задачи практического использования полученного уравнения. Концепция многомерного факторного пространства может быть полезна на всех этапах. Например, разработка *ротатабельных планов* второго порядка Бокса-Хантера была сопряжена с исследованием информационных контуров (поверхностей равной информации) в многомерном факторном пространстве. В задачах оптимизации исключительно эффективна идея движения по *поверхности отклика* в многомерном факторном пространстве.

Причина (*укр.* причина, польск. przyczyna < *чин, чинить. Лат.* causa) - *явление*, вызывающее возникновение другого явления. Познание *причинной связи* явлений играет огромную роль в интеллектуальном развитии человека (см. также *Следствие*).

Причина *ж.* начало, источник, вина, коренной повод действию, случаю; что производит последствия, что служит виною, рычагом, основной силой, начальным деятелем явления... **Причинность**, *лат.* via causalitatis, доведение до уверенности в чём либо, исходя от причины к причине... (В.И. Даль) (см. также *Следствие*).

Регрессионный анализ (от лат. regressio - обратное движение, отход, повторение) - раздел математической статистики, объединяющий практические методы исследования регрессионной зависимости между величинами по статистическим данным (см. *Регрессия*). Проблема регрессии в математической статистике характерна тем, что о распределении изучаемых величин **нет достаточной информации**. Цель регрессионного анализа состоит в проверке *статистических гипотез* о регрессии, а содержание заключается в получении уравнения регрессии, более или менее отражающего физическую сущность изучаемого явления, и вычислении статистических оценок неизвестных параметров, входящих в уравнение регрессии. Наибольшее распространение в науке и технике имеет регрессионный анализ парной зависимости вида $y=f(x)$ вследствие простоты анализа и наглядности в графической интерпретации. При этом величина X рассматривается как независимая переменная величина или *фактор*, а величина - Y как зависимая переменная величина или функция. Нередки случаи, когда выбор переменных (X,Y) или (Y,X) достаточно произволен с точки зрения формального описания, но с точки зрения регрессионного анализа это имеет большое значение (см. ниже). При изучении связи между двумя величинами по результатам наблюдений (x_1, y_1), (x_2, y_2),..., (x_n, y_n) в соответствии с теорией регрессии предполагается, что одна из них Y со значениями y имеет некоторое распределение вероятностей при фиксированном значении другой переменной x с некоторыми *математическим ожиданием* и *дисперсией*. Выбор модели регрессии определяется предположениями о форме зависимости $y=f(x)$. Для установления связей между величинами y и x в эксперименте используется модель, основанная на упрощённых, но правдоподобных допущениях: величина x является **контролируемой** величиной, значения которой заранее задаются при планировании эксперимента, а наблюдаемые значения Y представимы в виде

$$y_i=\varphi(x_i,\ \beta)+\varepsilon_i,\ i=1,\ 2,\dots,\ n, \qquad (1.42)$$

где величины ε_i характеризуют ошибки, независимые при различных измерениях и одинаково распределённые с нулевым средним и постоянной дисперсией σ^2. В действительности невозможно осуществить эксперимент, в котором отсутствовала бы ошибка измерения или поддержания на требуемом уровне независимой переменной X. Случай **неконтролируемой** переменной x (или переменная X со значениями x также имеет некоторое распределение вероятностей) отличается тем, что результаты наблюдений (x_1, y_1), (x_2, y_2),..., (x_n, y_n) представляют собой выборку из *совокупности* с некоторым двумерным распределением вероятностей. И в том и в другом случаях регрессионный анализ проводится одним и тем же способом, однако интерпретация результатов существенно различается (если обе исследуемые величины случайны, то *регрессионный анализ* в значительной степени дополняется методами корреляционного анализа).

Исследование регрессии, как правило, начинается с построения диаграммы рассеяния (называемого также корреляционным полем) точек (x_i, y_i). Во всех случаях по форме графика $y=\varphi(x)$ можно получить предварительное представление о силе зависимости y от x или об отсутствии оной. Визуальный анализ формы графика $y=\varphi(x)$ исключительно важен, и его значение невозможно переоценить. Например, если расположение этих точек на графике близко к прямолинейному, то допустимо использовать в качестве приближения линейную регрессию; если вид корреляционного поля ближе к параболе, гиперболе, следует попробовать полиномиальную модель; если зависимость имеет более сложный характер (например, имеются точки максимума и (или) минимума, участки стабилизации), то чисто статистический *регрессионный анализ* недостаточно эффективен, целесообразно подобрать детерминистическую модель, *отражающую* физическую сущность процесса. Стандартный метод оценки линии регрессии основан на использовании полиномиальной модели ($n \geqslant 1$):

$$y=b_0+b_1x+b_2x^2+b_3x^3+\ldots+b_nx^n. \qquad (1.43)$$

Оценки b_0, $b_1, \ldots,$ b_n принято осуществлять методом наименьших квадратов. Многочлен (1.43) характеризует так называемую эмпирическую линию регрессии, которая служит статистической оценкой **неизвестной истинной** линии регрессии.

Регрессионный анализ является одним из наиболее распространённых методов обработки результатов наблюдений при изучении зависимостей в биологии, химии, физике, экономике, технике, в различных технологиях и других областях. См. также *Регрессия*.

Регрессия (от лат. regressio - обратное движение, отход, повторение) в теории вероятностей и математической статистике - зависимость среднего значения какой-либо величины от некоторой другой величины или от нескольких величин. В отличие от функциональной зависимости $y=f(x)$, когда каждому значению независимой переменной x соответствует только одно значение величины y, при **регрессионной связи** одному и тому же значению x соответствует несколько значений величины y (конкретно, столько - сколько раз производилось наблюдение, измерение, определение и т.п. величины y при фиксированном значении независимой величины x). Если при каждом значении $x=x_1$ наблюдается n_i значений $y_{i.1}$, $y_{i.2}, \ldots,$ $y_{1.ni}$ величины y, то зависимость средних арифметических

$$\bar{y}_i = (y_{i.1}+y_{i.2}+\ldots+y_{1.ni})/n_1 \qquad (1.44)$$

от x_1 является **регрессией** в статистическом понимании этого термина. В пределе $n_i=1$. Методология регрессии основана на допущении "абсолютной" точности фиксации и "независимости" переменной X, на предположении, что случайные величины X и Y с заданным совместным распределением вероятностей связаны вероятностной зависимостью: при каждом фиксированном значении $X=x$ величина Y является случайной величиной с определённым (зависящим от значения x) условным распределением вероятностей. Уравнение $y=f(x)$, в котором x играет роль "независимой" переменной, называется **уравнением регрессии**, а соответствующий график - **линией** или **кривой регрессии** величины Y по X; переменная x называется **регрессионной переменной** или **регрессором**.

В практике экспериментальных исследований объём данных всегда ограничен, т.е. всегда есть дефицит информации о форме совместного распределения вероятностей, и соответственно более или менее реальна только задача нахождения уравнения ***приближённой*** регрессией. Основная задача исследователя при этом заключается в выборе из набора возможных функций $f(x)$, принадлежащих заданному классу, такой функции, которая минимизирует математическое ожидание $E(Y-f(x))^2$. Такая функция называется **средней квадратической регрессией**.

Первоначально термин "регрессия" был употреблён Ф. Гальтоном (1886) в теории наследственности в следующем специальном смысле: "возвратом к среднему состоянию" (regression to mediocrity) было названо явление, состоящее в том, что дети тех родителей, рост которых превышает среднее значение на a единиц, имеют в среднем рост, превышающий среднее значение меньше, чем на a единиц.

Ротатабельность (*англ.* rotatable - вращающийся, поворотный < *лат.* rotatio - круговое (вращательное) движение; rotatus - кружение, вращение) - *инвариантность* к *ортогональному* вращению *системы координат корреляционной матрицы* $(X^T X)^{-1}$ в задачах *планирования эксперимента*. Другими словами, ротатабельность - постоянство ошибки предсказания функции отклика $y_{расч}$ моделью процесса (уравнением регрессии) относительно истинного значения в пределах радиуса ядра плана, $0<\rho<1$. Ротатабельность - свойство плана эксперимента, позволяющее получить симметричные информационные контуры поверхности отклика в многомерном пространстве.

Свойство ротатабельности не инвариантно к изменению масштабности независимых переменных. См. также *Униформ-ротатабельное планирование*.

Следствие - то, что *физически* или логически с *необходимостью* вытекает из чего-либо другого, как из *причины* или основания.

Случайная величина - величина, значение которой невозможно предсказать исходя из условий эксперимента или наблюдений. Случайная величина - одно из основных понятий теории вероятностей. В самом общем смысле случайная величина - это некоторая переменная величина, принимающая в зависимости от случая те или иные значения с определёнными вероятностями. Так, например, число очков выпадающих на верхней грани игральной кости, представляет собой случайную величину, принимающую значения 1, 2, 3, 4, 5, 6 с вероятностями 1/6 каждое. Важнейшей характеристикой случайной величины служит её распределение вероятностей.

Случайные величины могут изменяться непрерывно (температура, давление, концентрация, скорость, пористость породы, радиус частиц дисперсной фазы) или дискретно (число "очков" в игре, число частиц, число дефектов, число отказов). Например, если представляет интерес количество монет, всё ещё находящихся в обращении, как функция их возраста, то можно реализовать два подхода к решению этой проблемы. Если в качестве случайной величины использовать год чеканки монеты, т.е. $X=\ldots,$ 1973, 1974, 1975,..., то анализировать придётся дискретные случайные величины, а если массу монеты, - то непрерывные случайные величины.

В науке и технологии результаты экспериментов, в основном, представляют собой непрерывные случайные величины. В некоторых случаях для случайной величины определяют законы распределения вероятностей: функцию распределения непрерывной случайной величины $P(X<x)=F(x)$ и плотность распределения вероятностей случайной величины X - $p(x)=F'(x)$. Необходимо различать распределение вероятностей собственно случайной величины и распределение вероятностей ошибки её определения. В большинстве случаев распределение вероятностей ошибок подчиняется закону нормального распределения. См. также *Математическое ожидание*, *Дисперсия*.

Совокупность - понятие теории статистического выборочного метода. В математической статистике совокупностью называется множество каких-либо однородных элементов, из которого по определённому правилу выделяется некоторое подмножество, называемое *выборкой*. Например, при приёмочном статистическом контроле в роли совокупности выступает множество всех изделий, подлежащее общей характеризации. В простейших случаях контролируемая выборка извлекается из совокупности случайно (наугад), что с точки зрения теории вероятностей означает: если совокупность содержит N элементов и отбирается выборка из n элементов ($n<N$), то выбор должен быть осуществлён таким образом, чтобы для любой группы из n элементов вероятность быть

извлечённой равнялась $n!(N-n)!/N!$.

В практике экспериментальных исследований и в математической статистике выборкой из совокупности принято также называть результаты измерений какой-либо физической величины, подверженной случайным ошибкам. В этом случае под совокупностью подразумеваются все возможные значения физической величины. Для решения практических задач бесконечное множество значений интереса не представляет; практический интерес представляют те или иные характеристики соответствующей функции распределения $F(x)$. В этом случае выборка из бесконечной совокупности представляет собой наблюдаемые значения нескольких случайных величин, по которым определяются необходимые параметры.

Состоятельность оценки - статистическая оценка параметра распределения вероятностей, обладающая тем свойством, что при увеличении числа наблюдений вероятность отклонений оценки от оцениваемого параметра на величину, превосходящую некоторое наперёд заданное число, стремится к нулю. Оценка параметра называется состоятельной, если по мере роста числа наблюдений $n \to \infty$ она стремится к математическому ожиданию оцениваемого параметра. Так, выборочное среднее и выборочная дисперсия представляют собой состоятельные оценки соответственно математического ожидания и дисперсии нормального распределения.

Стандартизация случайной величины - преобразование случайной величины с целью получения распределения с центром в нуле и *дисперсией*, равной единице. Стандартизация случайной величины производится путём вычитания из неё среднего значения выборки и деления на *квадратичное отклонение*. См. также *Величина параметрическая, Кодирование переменных, Нормализация, Нормирование переменных, Приведение переменных.*

Стандартное отклонение или **стандарт** - то же, что *квадратичное отклонение*. Стандартным отклонением в теории и практике экспериментальных исследований принято называть корень квадратный из выражений (1.5), (1.11), (1.40), (1.45), (1.65) и т.п. См. также *Дисперсия воспроизводимости.*

Статистических гипотез проверка (*греч.* υποθεσιζ - основание, принцип, предположение, **гипотеза**; υποτιθημι - полагать в основание, принимать что-либо за основание, предполагать) - один из основных разделов математической статистики, объединяющий методы проверки соответствия статистических данных некоторой статистической гипотезе (гипотезе о вероятностной природе данных). Проблема проверки гипотезы статистическими методами возникает в тех случаях, когда гипотезу нельзя проверить непосредственно и приходится довольствоваться проверкой некоторых следствий, которые логически вытекают из содержания гипотезы. Если следствия, вытекающие из предполагаемой гипотезы, невозможны или противоречат физической сущности процесса, значит гипотеза неверна. С другой стороны, если те или иные события невозможны или возможны с очень малой вероятностью, но всё-таки происходят, то гипотезу также приходится отвергать. Очевидно, что, придерживаясь подобной логики рассуждений, с некоторой вероятностью можно отвергнуть гипотезу, а выявить физическую сущность изучаемого явления или показать, что же происходит на самом деле, невозможно.

Процедуры проверки статистических гипотез позволяют принимать или отвергать статистические гипотезы, возникающие при обработке или интерпретации результатов наблюдений (результатов измерений переменных). **Правило**, в соответствии с которым принимается или отклоняется та или иная гипотеза, называется **статистическим критерием**. Проверка статистической гипотезы начинается с формулировки подходящей гипотезы об исследуемой переменной. Обычно такая гипотеза называется нулевой, обозначается H_0 и по существу является **гипотезой об отсутствии различия**. Например при оценке значимости (значимого отличия от нуля) коэффициентов уравнения регрессии подходящей нулевой гипотезой будет гипотеза о равенстве коэффициента уравнения регрессии нулю

$$H_0 : b_j = 0.$$

Подходящей альтернативной гипотезой в этом случае будет неравенство

$$H_1 : b_j \neq 0.$$

Некоторая неопределённость ответа на вопрос о равенстве коэффициента b_j нулю заключается в том, что при вычислении коэффициентов уравнения регрессии b_j по соответствующим формулам ноль получить практически невозможно. Коэффициент b_j может быть очень малым числом, но не нулём и при этом его величина будет соответствовать его сущности. И, наоборот, коэффициент b_j может быть весьма значителен по величине, но это будет противоречить физической сущности процесса. Проблема может быть решена, если коэффициент b_j сравнивать с его же квадратичным отклонением s_{bj}, которое соответственно может быть и очень малым числом, и очень большим, вопрос в соизмеримости b_j и s_{bj}. Поэтому, сохранив сущность, нулевую гипотезу сформулируем по другому и будем проверять гипотезу об отсутствии различия между коэффициентом b_j и его квадратичным отклонением s_{bj}. Если статистику критерия построить в виде отношения

$$t_{оп} = \frac{|b_j|}{s_{b_j}}, \qquad (1.45)$$

то, сравнивая опытный t-критерий с t-критерием, критическое значение которого получено независимым путём, можно принять или отвергнуть нулевую гипотезу.

Построение критерия определяется выбором подходящей функции $\theta = f(X_1, X_2, \ldots, X_n)$ от результатов набюдений $X_1, X_2, \ldots, X_n$, которая служит мерой расхождения между опытными и *гипотетическими* значениями. Функция θ является случайной величиной и называется **статистикой** критерия. Центральный момент при выборе функции θ заключается в том, что её теоретическое распределение вероятностей может быть вычислено **независимым** путём, при общем допущении, что проверяемая гипотеза верна и её распределение **не зависит** от характеристик гипотетического распределения. Распределение статистики θ табулируется для различных *чисел степеней свободы* ν и *уровней значимости* α. С помощью этого распределения находится критическое значение θ^α такое, что если гипотеза верна, то вероятность события $\theta > \theta^\alpha$ равна α (рис. 1.1). Область значений $(x_1, x_2, \ldots, x_n)$, для которых $\theta(x_1, x_2, \ldots, x_n) > \theta^\alpha$, называется **критической областью**. Это область отклонения гипотезы H_0. Если в конкретном случае обнаружится, что $\theta > \theta^\alpha$, то гипотеза отвергается; при этом считается, что *значимость* этого расхождения равна α, а *доверительная вероятность* правильности отклонения гипотезы равна $P = 1 - \alpha$. Если в конкретном случае $\theta < \theta^\alpha$, то считается, что с вероятностью P гипотеза верна. Такого рода критерии используются как для проверки параметров распределения на *значимость*, так и для проверки гипотез о самих распределениях.

Например, пусть имеется гипотеза H_0 о физической сущности некоторого *процесса* $Y_1, Y_2, Y_3, \ldots$. Для подтверждения этой гипотезы необходимо определить параметры математической модели, соответствующей природе изучаемого процесса, и проверить математическую модель

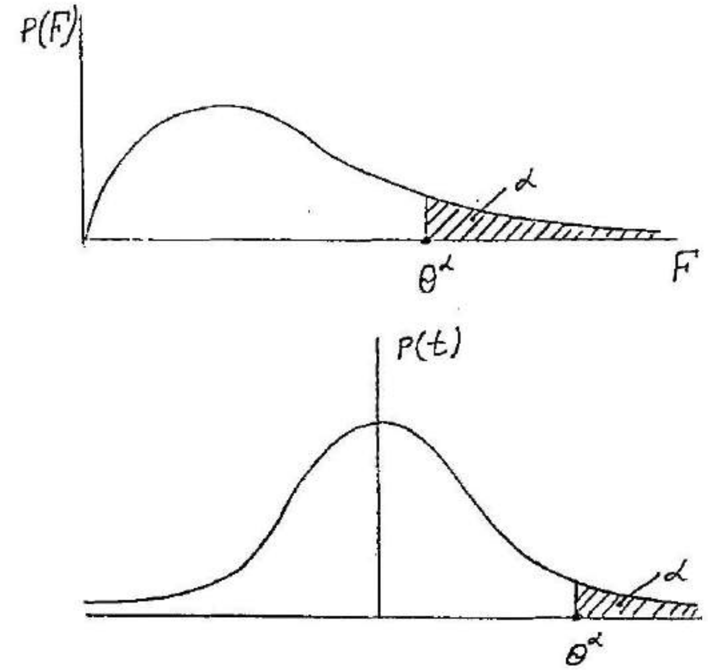


Рис. 1.1. Типичные F- и t-распределения. Заштрихована критическая область, составляющая α площади под кривой (общая площадь под кривой равна 1)

на *адекватность*. В общем случае в результате экспериментов необходимо получить две *выборки*: выборку размерности n для определения собственно параметров математической модели y_1, $y_2, \ldots, y_n$ и выборку размерности $n_{оп}$ для оценки точности эксперимента y_1, $y_2, \ldots, y_{n_{оп}}$:

$$s^2_{оп} = \frac{1}{n_{оп}-1} \cdot \sum_{i=1}^{n_{оп}} (y_i - \bar{y})^2; \qquad (1.46)$$

$$\bar{y} = \frac{1}{n_{оп}} \cdot \sum_{i=1}^{n_{оп}} y_i, \qquad (1.47)$$

где $s^2_{оп}$ - *несмещённая оценка* генеральной дисперсии или выборочная дисперсия (1.5). В качестве нулевой гипотезы будем рассматривать гипотезу об отсутствии различия опытной дисперсии и дисперсии адекватности. Статистический критерий для проверки гипотезы H_0 можно построить следующим образом:

$$F^{оп}_{\nu_1, \nu_2} = \frac{S^2_{ад}}{S^2_{оп}}; \qquad (1.48)$$

$$S^2_{ад} = \frac{1}{n-1} \cdot \sum_{i=1}^{n} (y_i - \hat{y}_i)^2, \qquad (1.49)$$

где F – критерий Фишера; $s^2_{ад}$ – *дисперсия адекватности*; ν_1 – *число степеней свободы* числителя (в данном случае *дисперсии адекватности*); ν_2 – число степеней свободы знаменателя (в данном случае *дисперсии воспроизводимости*); $y_{1.расч}$ – значения функции отклика, рассчитанные по проверяемой модели с параметрами, определёнными по выборке y_1, $y_2, \ldots,$ y_n, причём каждому $y_{1.расч}$ соответствует $y_{1.эксп}$. Опытный критерий Фишера сравнивается с табличным значением, которое получается независимо с помощью формулы плотности распределения критерия Фишера:

$$p(F) = \frac{\Gamma\left(\frac{\nu_1+\nu_2}{2}\right)}{\Gamma\left(\frac{\nu_1}{2}\right)\cdot\Gamma\left(\frac{\nu_2}{2}\right)} \cdot \frac{\nu_1^{\nu_2/2}\ \nu_2^{\nu_1/2}\ F^{\nu_1/2-1}}{(\nu_1 F+\nu_2)^{(\nu_1+\nu_2)/2}}. \qquad (1.50)$$

Если $F^{оп} \leqslant F^{\alpha}$, то мы оказываемся в области принятия нулевой гипотезы, т.е. математическая модель с уровнем значимости α адекватна; если $F^{оп} > F^{\alpha}$, то мы попадаем в критическую область, область отклонения нулевой гипотезы – модель неадекватна. Если $F^{оп} \leqslant F^{\alpha}$, то это ещё не означает действительного подтверждения гипотезы, так как невысокая точность экспериментальной работы приводит к завышению дисперсии воспроизводимости и, соответственно, к занижению опытного критерия Фишера; к такому же результату приводит уменьшение числа опытов на воспроизводимость.

При решении вопроса о принятии или отклонении какой-либо гипотезы H_0 с помощью любого критерия, основанного на результатах наблюдения, могут быть допущены ошибки двух родов. Ошибка первого рода совершается тогда, когда отвергается верная гипотеза H_0. Ошибка второго рода совершается в том случае, когда гипотеза H_0 принимается, а на самом деле верна не она, а какая-либо альтернативная гипотеза H_1. Вероятность допустить ошибку первого рода равна α, т.е. уровню значимости критерия; вероятность допустить ошибку второго рода равна P, т.е. доверительной вероятности. Эти ошибки не равноценны.

Стьюдента критерий, t-критерий – *критерий значимости*, основанный на распределении Стьюдента и используемый для проверки гипотез о средних значениях нормальных распределений. Пусть результаты наблюдений x_1, $x_2, \ldots,$ x_n – взаимно независимые нормально распределённые случайные величины с **неизвестными параметрами** M_x и $б^2$, т.е. $N(M_x,\ б^2)$. При отсутствии грубых и систематических ошибок результат первоначальной обработки наблюдений x_{cp} совпадает с математическим ожиданием M_x с большей или меньшей точностью, зависящей от объёма выборки n:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i; \tag{1.51}$$

$$s^2{}_x = \frac{1}{n-1} \cdot \sum_{i=1}^{n} (x_i - \bar{x})^2, \tag{1.52}$$

где x_{cp} - оценка математического ожидания M_x, а $s^2{}_x$ - оценка генеральной дисперсии $б^2$. В тех случаях, когда $s_x \approx |x_{cp}|$ или $s_x \gg |x_{cp}|$, говорят, что "оценка x_{cp} математического ожидания M_x *незначимо* отличается от нуля"; в этом случае может быть два исхода: либо $M_x=0$, либо $M_x \neq 0$, в обоих случаях необходимо повысить точность проведения наблюдений и (или) эксперимента и (или) увеличить объём выборки. В тех случаях, когда $s_x \ll |x_{cp}|$ и возникает задача собственно оценки точности наблюдений, а более строго, определения интервала значений, с той или иной вероятностью "накрывающего" неизвестное значение параметра исследуемого распределения. Эта задача была решена в 1908 г. У. Госсетом, известным под псевдонимом Student:

$$\bar{x} - \frac{s_x}{\sqrt{n}} \cdot t \leqslant M_x \leqslant \bar{x} + \frac{s_x}{\sqrt{n}} \cdot t, \tag{1.53}$$

где t - случайная величина, зависящая только от *числа степеней свободы* $\boldsymbol{\nu}$ выборочной дисперсии и *уровня значимости* $\boldsymbol{\alpha}$. Случайная величина t имеет распределение Стьюдента или t-распределение:

$$t_{n-1} = \frac{(\bar{x} - M_x)}{s_x} \sqrt{n}, \tag{1.54}$$

Плотность распределения вероятностей $p(t)$ этой случайной величины имеет вид

$$p(t) = \frac{1}{\sqrt{\pi\nu}} \cdot \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)} \cdot \left(1 + \frac{t^2}{\nu}\right)^{\frac{\nu+1}{2}}, \tag{1.55}$$

где $\Gamma(\nu)$ - гамма-функция; ν - число степеней свободы выборки; t - критерий Стьюдента, $-\infty<t<+\infty$. Если среднее $x_{\text{ср}}$ и выборочная дисперсия s^2_x определяются по одной и той же выборке, то $\nu=n-1$. Формула (1.55) использована для табулирования критерия Стьюдента при различных уровнях значимости α. Формулой (1.53) пользуются практически при обработке экспериментальных данных для определения *доверительного интервала* $x_{\text{ср}}$. Очевидно, что выражение

$$\frac{s_x}{\sqrt{n}}\cdot t^{\alpha}_{\nu}, \tag{1.56}$$

является доверительным интервалом для среднего значения выборки. См. также *Статистических гипотез проверка*.

Униформ-ротатабельное планирование {*англ.* uniform - равномерный, однородный, единообразный < *лат.* uniformis - однообразный, простой; *англ.* rotatable - вращающийся, поворотный < *лат.* rotatio - круговое (вращательное) движение; rotatus - кружение, вращение; *лат.* planum - равнина, плоскость; planus - плоский, ровный; ясный, понятный)} - *планирование эксперимента*, у которого корреляционная матрица $(X^TX)^{-1}$ инвариантна к ортогональному вращению системы координат. Это условие удовлетворяется, если моменты информационной матрицы удовлетворяют соотношениям

$$\sum_{i=1}^{n} z^2_{i,j} = n\lambda_2, \quad j=1,\ 2,\dots,\ k; \tag{1.57}$$

$$\sum_{i=1}^{n} z^4_{i,j} = 3\cdot\sum_{i=1}^{n} z^2_{i,l}z^2_{i,m} = 3n\lambda_4; \quad l,\ m=1,\ 2,\dots,\ k;\ l\neq m, \tag{1.58}$$

где $z_{i,j}$ - *кодированные переменные*; λ_2 и λ_4 - произвольно выбираемые константы, удовлетворяющие неравенству

$$\frac{\lambda_4}{\lambda_2^2} > \frac{k}{k+2}, \tag{1.59}$$

которое является условием невырожденности матрицы $\boldsymbol{Z}^T\boldsymbol{Z}$. Остальные моменты информационной матрицы равны нулю. *Далее по тексту разд.2.2.*

Уровень значимости статистического критерия - вероятность ошибочно отвергнуть основную проверяемую гипотезу, когда она верна. Понятие "уровень значимости" возникло в связи с задачей проверки согласованности теории с опытными данными. Если, например, в результате наблюдений регистрируются значения n случайных величин X_1, $X_2, \ldots, X_n$ и требуется по этим данным проверить гипотезу H_0, согласно которой совместное распределение величин $X_1, X_2, \ldots, X_n$ обладает некоторым определённым свойством, то соответствующий статистический критерий **конструируется** с помощью подходящим образом подобранной функции $\theta=f(X_1, X_2, \ldots, X_n)$. Эта функция обычно принимает малые значения, когда гипотеза H_0 верна, и большие значения, когда H_0 ложна; такую гипотезу ещё называют **гипотезой об отсутствии различия** или **нулевой гипотезой**. Соответствующий *критерий значимости* представляет собой правило, согласно которому значимыми считаются значения θ, превосходящие некоторое критическое значение θ^α. В свою очередь выбор величины θ^α определяется заданным уровнем значимости α, который в случае отклонения гипотезы H_0 совпадает с вероятностью события $\{\theta>\theta^\alpha\}$. Центральный момент при проверке гипотезы H_0 заключается в том, что уровнем значимости α задаются до анализа выборки на основании физической сущности задачи и последствий от ошибочного принятия решения. Диапазон значений уровней значимости, принимаемых в науке и технике, достаточно широк: 0,1; 0,05; 0,02; 0,01; 0,001. Наиболее употребительно значение α=0,05. В теории статистической проверки гипотез уровень значимости называется *вероятностью ошибки первого рода*. Вероятность такой ошибки не больше **принятого** уровня значимости. Например, при α=0,05 можно совершить ошибку первого рода в пяти случаях из ста. Принятие основной проверяемой гипотезы, когда она неверна, называется *ошибкой второго рода*. Фиксация уровня значимости находится целиком в компетенции исследователя: он должен решать, какой риск при отклонении истинной гипотезы является допустимым.

В геологии обычно имеют дело с обстоятельствами большой неопределённости, например, объём образцов (кернов), извлекаемых из скважин при разведочном бурении, несоизмеримо меньше объёма исследуемой залежи. Нулевой гипотезой в данном случае будет являться гипотеза об отсутствии различия образцов исследуемой залежи от пустой породы; подтверждение гипотезы будет означать бесперспективность дальнейшего бурения, а опровержение - наличие той или иной нефтегазоносности. Если позволить допустить ошибку в одном случае из ста (α=0,01) или даже в одном случае из двадцати (α=0,05), то, имея в распоряжении керны из нескольких скважин, будет трудно отвергнуть нулевую гипотезу (т.е., доказать перспективность залежи), и возникнет необходимость во всё большем и большем объёме образцов, получить которые непросто. Принимая более скромные уровни значимости (α>0,1), можно быстрее прийти к заключению, хотя вероятность получить ошибочные выводы может оказаться очень высокой в сравнении со стандартами, принятыми в других областях.

Критерий значимости, с помощью которого гипотеза проверяется, конструируется таким образом, чтобы критическое значение критерия могло быть вычислено независимым путём в предположении, что проверяемая гипотеза верна. При этом значения собственно критерия располагаются вдоль оси абсцисс, функция представляет собой некоторую кривую, площадь под которой равна единице, а критерий значимости α точно равен площади под кривой в критической области, т.е. в области отклонения проверяемой гипотезы. Примерный характер кривой приведён на рис. 1.1, но в общем случае он может быть другим. Представим себе, что нефтяная компания сконструировала статистический критерий прогноза нефтегазоносности θ, состоящий из некоторых количественных переменных, позволяющих определять приоритеты при бурении. Цель применения статистического критерия - сделать более или менее верный прогноз продуктивности скважин. Нулевая гипотеза H_0 состоит в том, что керны отобраны из совокупности бесперспективных разрезов, альтернативная H_1 - что керны отобраны из нефтяного или газового месторождения. Альтернативных гипотез может быть несколько.

Если принять уровень значимости, например α=0,05 (рис. 1.2, а), то очень малая часть образцов окажется отличающейся от неперспективной породы (нулевой совокупности). Если же окажется, что образцы отличаются от неё, то это почти наверняка даст открытие месторождения при бурении. Компания получит очень высокое соотношение для числа успехов при бурении, но при этом пропустит много залежей, которые могли бы оказаться продуктивными. Другими словами, компания будет редко бурить, редко совершать ошибки первого рода, но, соответственно, оставит много месторождений неоткрытыми.

Если принять уровень значимости побольше, например α=0,40 (рис. 1.2, б), то придётся осуществлять частую сеть бурения, соответственно частота неперспективных скважин будет значительно выше, но вероятность пропуска залежи будет меньше. При такой практике принятия решений компания будет часто бурить, часто ошибаться, но и

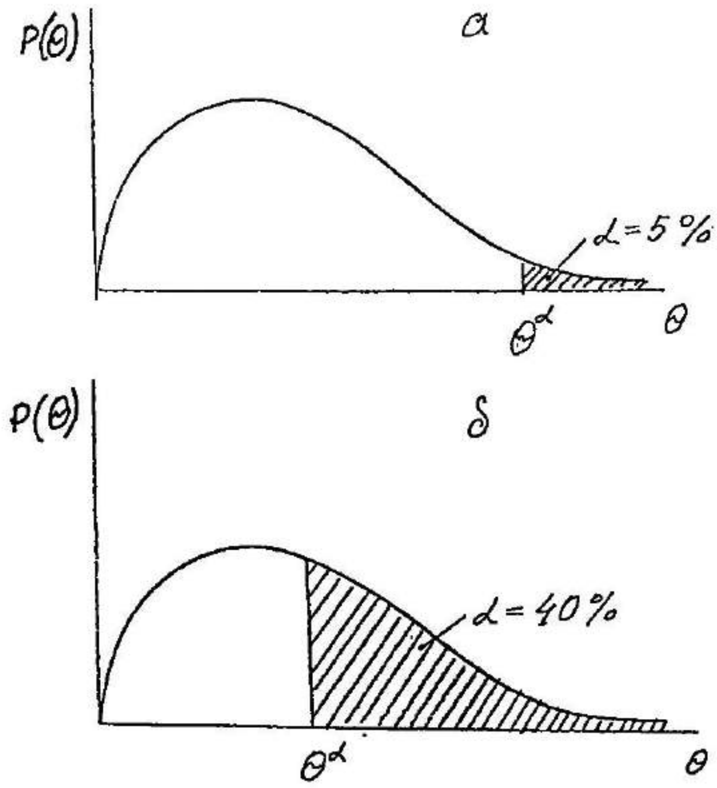


Рис. 1.2. Распределение статистики гипотетического критерия θ с критической областью α отклонения гипотезы о бесперспективности бурения

значительно меньше месторождений нефти останется неоткрытой.

В нефтяной промышленности случаи получения отрицательного результата при бурении перспективных площадей встречаются значительно чаще, чем последствия получения положительного результата при бурении пустых скважин. Причина этого состоит в том, что финансовый успех одного большого открытия может покрыть стоимость нескольких десятков пустых скважин. Оценка вероятности успеха одного из методов бурения в нефтяной промышленности США примерно равна 10%. Если бы эти скважины были пробурены на основе применения статистических критериев, то эта оценка соответствовала бы уровню значимости примерно α=0,9 [16].

Рассмотренный выше пример иллюстрирует выбор уровня значимости так называемого одностороннего критерия, поскольку нефть либо есть в залежи, либо нет, т.е. соответствующий критерий располагается только в положительной области. В тех случаях, когда физическая величина может принимать значения как положительные, так и отрицательные, либо критерий, соответствующий нулевой гипотезе, может располагаться в обеих областях декартовой системы координат, применяют двусторонний критерий значимости. Термин "нулевая гипотеза" возник оттого, что математическое ожидание критерия, соответствующее подтверждению основной проверяемой гипотезы, равно нулю (рис. 1.3). Очевидно, что вероятность попадания критерия как в левую область, так и в правую одинакова, соответственно одинакова и вероятность попадания критерия Θ как в левую, так и в правую критические области. Таким образом, уровень значимости α, т.е. вероятность отклонения нулевой гипотезы, распадается на две равные области, каждая из которых равна $\alpha/2$. Критическое значение Θ при этом увеличивается. Из этого следует важный вывод - приняв уровень значимости α для проверки гипотезы H_0, табличное значение критерия Θ следует брать для вдвое меньшего значения, т.е. для $\alpha/2$. Соответствующие значения критериев иногда записывют, например, так: для левой критической области $\Theta^{1-\alpha/2}$, для правой - $\Theta^{\alpha/2}$ [2].

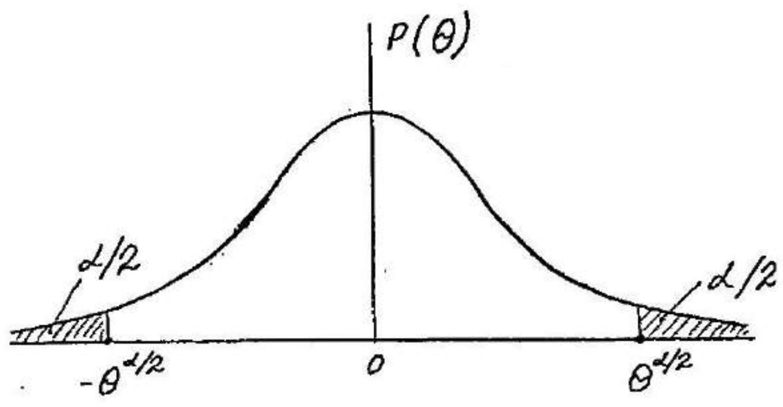


Рис. 1.3. Левая $-\Theta^{\alpha/2}$ и правая $\Theta^{\alpha/2}$ критические области гипотетического двустороннего критерия Θ^{α}

При выборе уровня значимости следует учитывать ущерб, неизбежно возникающий при использовании любого критерия значимости. Так, например, если уровень значимости чрезмерно велик, то основной ущерб будет происходить от ошибочного отклонения правильной гипотезы; если же уровень значимости мал, то ущерб будет, как правило, возникать от ошибочного принятия гипотезы, когда она ложна. Пример Дж.С.Дэвиса - крайний случай, но он показывает, что исследователь должен принимать решения на границе области риска и выбирать уровень значимости в соответствии с конкретными обстоятельствами.

Фактор (от лат. factor - мастер, создатель, виновник; facto - делать, совершать) - 1) движущая сила, причина какого-либо процесса, явления; 2) существенное обстоятельство в каком-либо процессе, явлении; 3) независимая переменная физическая величина или *аргумент*. При наличии трёх и более аргументов принято говорить о пространстве независимых переменных или *факторном пространстве*.

Фишера критерий, *F*-критерий - случайная величина, имеющая плотность распределения вероятности

$$p(F) = \frac{\Gamma\left(\frac{\nu_1+\nu_2}{2}\right)}{\Gamma\left(\frac{\nu_1}{2}\right)\cdot\Gamma\left(\frac{\nu_2}{2}\right)} \cdot \frac{\nu_1^{\nu_2/2}\,\nu_2^{\nu_1/2}\,F^{\nu_1/2-1}}{(\nu_1 F+\nu_2)^{(\nu_1+\nu_2)/2}} \qquad (1.60)$$

и зависящая только от *чисел степеней свободы* ν_1 и ν_2; Γ - гамма-функция. Критерий Фишера применяется для проверки гипотез об *однородности выборочных дисперсий*, например $s^2{}_1$ и $s^2{}_2$. *Нулевая гипотеза* формулируется как гипотеза об отсутствии различия дисперсий, $H_0: s^2{}_1=s^2{}_2$, альтернативная - дисперсии $s^2{}_1$ и $s^2{}_2$ неоднородны и характеризуют выборки из разных совокупностей, $H_0: s^2{}_1=s^2{}_2$. Опытный критерий Фишера конструируется в виде отношения

$$F^{\text{оп}} = \frac{s^2{}_1/\sigma^2{}_1}{s^2{}_2/\sigma^2{}_2}, \qquad (1.61)$$

где $\sigma^2{}_1$ и $\sigma^2{}_2$ - неизвестные генеральные дисперсии первой и второй совокупностей. В условиях нулевой гипотезы генеральные дисперсии равны ($\sigma^2{}_1=\sigma^2{}_2$) выборочные дисперсии однородны и отношение $s^2{}_1/s^2{}_2$ может непосредственно использоваться для проверки гипотезы. Опытный критерий Фишера

$$F^{\text{оп}} = \frac{s^2{}_1}{s^2{}_2} \Leftrightarrow F^{\alpha}_{\nu_1,\,\nu_2}, \qquad (1.62)$$

сравнивается с табличным значением для чисел степеней свободы числителя ν_1, знаменателя ν_2 и *уровня значимости* α, причём в числителе должна быть большая дисперсия. В этом случае табличное значение критерия Фишера является предельным соотношением однородных дисперсий (см. также *Статистических гипотез проверка*).

Функция (от лат. functio - исполнение, совершение, служебная обязанность, функция)) - явление, зависящее от другого и изменяющееся по мере изменения этого другого явления.

Функция - одно из основных понятий математики, выражающее зависимость одних переменных *величин* от других. Слово "величина" в этом определении функции понимается в самом широком смысле: именованное число, отвлечённое число (действительное или комплексное), несколько чисел (т.е. *точка* пространства) и, вообще, *элемент* любого множества. В простейшем случае действительной функцией одной действительной переменной величины, когда *величина* - действительное число, понятие функции определяется следующим образом. Пусть каждому числу x из заданного множества X поставлено в соответствие число y, обозначаемое $y=f(x)$; в этом случае говорят, что на множестве X задана *функция*

$$y=f(x),\quad x\in X,$$

где x - независимая переменная величина, или *аргумент*, или *фактор*; y - зависимая переменная величина, или *функция*; X - множество значений, которые может принимать x, - *область определения*, или *область задания* функции. Выражение "поставлено в соответствие" означает, что указан определённый способ, по которому для каждого $x\in X$ находится $y=f(x)$. Функция может быть задана различными способами: аналитически, графически, таблично и в словесной форме.

Аналитический способ наиболее распространён, при этом функция задаётся формулой, указывающей какие вычислительные операции необходимо произвести над x, чтобы получить y. Например, $y=3+\ln x$; $y=3+4x+5x^2+6x^3$.

При табличном способе задания функция задаётся в виде таблицы, в которой каждому значению аргумента указывается соответствующее ему значение функции. Такой способ задания функции является результатом подавляющего большинства экспериментальный исследований и именно таблично заданная функция используется для статистического анализа *данных*, *восстановления зависимости* и *проверки гипотез* об исследуемом процессе. Широко распространена и обратная процедура, когда аналитическая функция *табулируется*, т.е. представляется в виде таблиц: таблиц логарифмов, тригонометрических функций, а такие функции как распределения Стьюдента, Фишера и др. практически используются только в виде таблиц.

Графический способ необходим исследователю для визуального представления и первичного анализа экспериментально полученной зависимости. Графиком функции $y=f(x)$ называется множество точек плоскости с прямоугольными координатами (x, y), где $x \in X$. Однако функция, заданная графически, недостаточно определена с чисто математической точки зрения и её табличное представление используется в качестве задачи *восстановления зависимости*.

Действительная функция нескольких действительных переменных широко используется при математическом *моделировании* технологических процессов. В случае множества *аргументов (факторов)* принято говорить о *функции отклика* и *факторном пространстве* $y=f(x_1, x_2, \ldots, x_k)$, где k - количество независимых переменных (*факторов*) или размерность факторного пространства. При статистическом моделировании размерность факторного пространства не ограничена, а при построении *детерминистических моделей* обычно ограничиваются четырьмя факторами - три координаты нашего трёхмерного мира и время (например процесс нестационарного теплообмена, массообмена и др.).

Впервые термин "функция" использовал Г.В. Лейбниц в рукописи 1673 г., оставшейся неизданной, под названием "*Обратный метод касательных или* [рассуждения] *по поводу функций*" по поводу задачи, в которой требовалось определить координаты, исходя из заданного свойства касательных к кривой. Лейбниц называл функцией любую линию (длина которой зависит от положения некоторой точки на данной кривой), которая в общепринятом смысле слова выполняет свою *функцию*" в фигуре: иначе говоря, играет роль касательной, нормали, подкасательной и т.д. и которая таким образом "функционирует". Подобное соглашение о смысле слова "функция" было принято и в некоторых других его статьях, опубликованных в 1692 и 1694 гг., и в том же смысле это слово появилось в 1697 г. в работе Иоганна Бернулли.

Функция отклика - *функция* нескольких действительных переменных, используемая при математическом *моделировании* технологических процессов. Отличительной особенностью функции отклика является множество её значений $y_1, y_2, \ldots, y_m$ для каждой конкретной точки *факторного пространства* $(x_1, x_2, \ldots, x_k)_i$ при экспериментальном определении, где k - количество независимых переменных (*факторов*) или размерность факторного пространства. При этом $y_1, y_2, \ldots, y_m$ рассматриваются как *случайные величины*, имеющие то или иное *распределение вероятностей*. Если в результате обработки экспериментальных *данных* получено *уравнение регрессии* $y=f(x_1, x_2, \ldots, x_k)$, то y также называют функцией отклика, поскольку между найденной зависимостью $y=f(x_1, x_2, \ldots, x_k)$ и пространством $\boldsymbol{Y}=f(\boldsymbol{X})$ может быть большая или меньшая степень соответствия, вплоть до несоответствия.

В равной степени термин "функция отклика" используется и для парных зависимостей $y=f(x)$, определяемых по результатам наблюдений. В отличие от функциональной зависимости $y=f(x)$, когда каждому значению независимой переменной x соответствует только одно значение величины y, при **регрессионной связи** одному и тому же значению x соответствует несколько значений величины y (конкретно столько, сколько раз производилось наблюдение, измерение, определение и т.п. величины y при фиксированном значении независимой величины x). См. также *Адекватность*, *Значимости уровень*, *Статистических гипотез проверка*.

Число степеней свободы характеризует информационный потенциал выборки. Это всегда целое положительное число, равное разности между числом наблюдений в выборке и числом параметров, определённых по данным выборки. Число степеней свободы обозначается греческой буквой $\boldsymbol{\nu}$, а число параметров, определённых по выборке, - латинской буквой l; таким образом, $\nu=n-l$. Число параметров, определённых по выборке ещё называется *числом связей*, наложенных на выборку. Так вот, с точки зрения теории информатики число степеней свободы равно числу параметров, которое ещё можно определить по выборке после той или иной обработки, а с точки зрения математической статистики $\boldsymbol{\nu}$ равно числу независимых источников информации, по которым вычисляется тот или иной выборочный параметр. Дело в том, что используя одну и ту же выборку, невозможно решить сразу две задачи: оценить параметры совокупности и применить соответствующий критерий для проверки достоверности полученных оценок без какой-либо компенсации, связанной с двукратным обращением к имеющемуся массиву наблюдений. Такой компенсацией является уменьшение знаменателя в формуле выборочной дисперсии от числа наблюдений n до числа независимых источников информации оцениваемого параметра $\boldsymbol{\nu}$. Если, например, математическое ожидание M_x оценивается по результатам пяти независимых наблюдений

$$\bar{x} = (x_1+x_2+x_3+x_4+x_5)/5; \qquad (1.63)$$

то результат имеет пять степеней свободы. Дисперсия оценивается по пяти квадратам разностей $(x_i-x_{cp})^2$, однако независимо вычисляются только четыре из этих разностей, так как определив четыре, пятую уже можно вычислить следующим образом:

$$x_5-\bar{x} = 5\bar{x}-(x_1+x_2+x_3+x_4). \qquad (1.64)$$

Поэтому имеется только четыре независимых источника информации, по которым вычисляется выборочная дисперсия. Бывают случаи, когда в качестве оценки математического ожидания M_x используется величина, не зависимая от рассматриваемой выборки (например m_x), т.е. оценка определяется независимым путём. В таких случаях для выборочной дисперсии следует пользоваться формулой (1.39):

$$s_x^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i-m_x)^2. \qquad (1.65)$$

Эффективность оценки - свойство оценки иметь больший или меньший доверительный интервал. Оценка параметра называется эффективной, если среди нескольких оценок того же параметра она обладает наименьшей дисперсией.

Эффект линейный, парного взаимодействия и квадратичный - коэффициент, соответственно, b_j, b_{lm} и b_{jj} уравнения регрессии. Термин "эффект взаимодействия" используется при обсуждении *корреляции* коэффициентов b_j, b_{lm} и b_{jj}, причём рассматриваются различные комбинации корреляции. В пассивном эксперименте рассматривается корреляция коэффициентов b_j между собой и свободным членом b_0. При обсуждении *униформ-ротатабельного планирования* рассматривается корреляция коэффициентов при парных произведениях $b_{l,u}$, $b_{l,m}$ и корреляция коэффициентов при квадратичных членах b_{ll}, b_{mm}. Зависимость коэффициента корреляции от угла поворота системы координат в пространстве имеет сложный характер (см.рис.2.5, а, б). См. также *Корреляция*.

Ядро плана второго порядка - план *полного факторного эксперимента* (план типа 2^k), который Д.Бокс и К.Уилсон предложили взять за основу для добавления опытов с целью получения уравнения регрессии второго порядка. Вокруг ядра плана второго порядка достраиваются специальным образом расположенные так называемые *звёздные точки* и добавляются точки в центре плана. Такие планы получили название *центральных композиционных планов*. Так, в матрице (2.1) первые четыре строки представляют собой ядро плана второго порядка, а в матрице (2.36) - первые восемь строк. См. также *Кодирование переменных, Ортогональное центральное композиционное планирование. Ротатабельность, Униформ-ротатабельной планирование.*

1.2. Метод наименьших квадратов в пространстве

При использовании методов планирования эксперимента вначале ограничиваются небольшим участком поверхности, для которого может быть справедливо линейное приближение. Соответственно, для описания этой поверхности можно использовать уравнение *линейной множественной регрессии*

$$y=b_0+b_1x_1+b_2x_2+\ldots+b_kx_k, \tag{1.66}$$

коэффициенты b_0, b_1, $b_2,\ldots$, b_k которого обычно находят методом *наименьших квадратов* (НК). Если в выбранной области кривизна поверхности относительно велика, то используется квадратное уравнение

$$y=b_0+b_1x_1+b_2x_2+\ldots+b_kx_k+b_{12}x_1x_2+b_{23}x_2x_3+\ldots+b_{k-1,k}x_{k-1}x_k+$$
$$b_{11}x_1^2+b_{22}x_2^2+\ldots+b_{kk}x_k^2. \tag{1.67}$$

где b_0, $b_1,\ldots$, b_k, b_{12}, $b_{23},\ldots$, $b_{k-1,k}$, b_{11}, $b_{22},\ldots$, b_{kk} - искомые коэффициенты уравнения регрессии, а, точнее, оценки соответствующих генеральных коэффициентов β_j, $\beta_{l,m}$ и β_{jj}.

С точки зрения ***метода*** НК последнее уравнение является также уравнением линейной множественной регрессии, поэтому прежде чем переходить к дальнейшему, необходимо обсудить смысл прилагательного "линейный" в рассматриваемой модели линейной регрессии $y=f(x_1, x_2,\ldots, x_k)$. См. разд. 1.1 "*Линейности условие метода НК*".

Метод НК получил своё название из-за его связи с минимизацией сумм квадратов вида

$$\Phi=\sum_{i=1}^{n}\left(y_i-f(x_1,\ x_2,\ldots,\ x_k)_i\right)^2, \tag{1.68}$$

где в рассматриваемом случае

$$f(x_1,\ x_2,\ldots,\ x_k)_i=\sum_{j=0}^{k}b_jx_{i,j}. \tag{1.69}$$

Как общий принцип он формулируется следующим образом: в качестве оценки вектора неизвестных параметров b_0, $b_1, \ldots,$ b_k в некотором выражении $\boldsymbol{\varphi}(\boldsymbol{X}, \boldsymbol{Y}, \boldsymbol{B})=0$, где $\boldsymbol{X}$ и $\boldsymbol{Y}$ - результаты наблюдений, а $\boldsymbol{B}$ - вектор параметров, следует взять такое $\boldsymbol{B}_{расч}$, которое обращает в минимум сумму (1.68). Так вот, если принять, что оценки неизвестных параметров распределения являются ***линейными*** функциями от результатов наблюдений, то метод НК даже при малых выборках обладает свойством оптимальности, состоящим в том, что он даёт несмещённые оценки, имеющие минимальную дисперсию. В случае нахождения центра распределения случайной величины (математического ожидания M_x) такой оценкой является арифметическое среднее (линейная функция от x_i)

$$M_x = \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i, \tag{1.70}$$

а оценкой генеральной дисперсии является выборочная дисперсия

$$s^2{}_x = \frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2. \tag{1.71}$$

Случаем также одного неизвестного параметра является парная зависимость вида $y=bx$, параметр b которой по методу НК находится из следующего требования:

$$\Phi = \sum_{i=1}^{n}\left(y_i - bx_i\right)^2 = \min. \tag{1.72}$$

Случаем двух неизвестных параметров также является парная зависимость вида $y=b_0+b_1x$, линейная и с точки зрения здравого смысла, и с точки зрения метода НК, параметры b_0 и b_1 которой по методу НК находятся аналогично:

$$\Phi = \sum_{i=1}^{n}\left(y_i - (b_0+b_1x)\right)^2 = \min. \tag{1.73}$$

Собственно парных зависимостей в природе достаточно мало, поскольку каждое событие в мире является результатом одновременного воздействия некоторого множества детерминированных факторов. Правильнее сказать, что классическая парная зависимость $y=f(x)$ является частным случаем множественной зависимости $y=f(x_1, x_2, \ldots, x_k)$, тем или иным *сечением*. Поэтому следующим уравнением, описывающим некоторые процессы в природе, является уравнение линейной множественной регрессии

$$y=b_0+b_1x_1+b_2x_2+b_3x_3+\ldots+b_kx_k. \qquad (1.74)$$

Человеку, живущему в четырёхмерном пространстве-времени и привыкшему строить соответствующие мысленные модели, трудно создавать и анализировать мысленную модель, образ многомерного пространства. Человеку, может быть, не проще, но эффективнее анализировать отдельные зависимости $y=f(x_1)$, $y=f(x_2),\ldots$, $y=f(x_k)$, обобщать и делать выводы впоследствии. Именно поэтому основной прогресс в развитии методов поиска параметров уравнения вида (1.74) связан с созданием и развитием вычислительных машин, когда проблема мысленного анализа многомерного пространства была заменена проблемами, например, корреляции коэффициентов b_j, вырожденности информационной матрицы (см. ниже) и т.п.

В общем случае модель, линейную относительно параметров b_j, вида (1.74) можно представить в виде

$$\boldsymbol{Y} = \boldsymbol{BX} + \varepsilon, \qquad (1.75)$$

где $\boldsymbol{Y}$ - вектор-столбец наблюдений размерности n; $\boldsymbol{X}$ - матрица независимых переменных (факторов) размерности $\{n\times(k+1)\}$; $\boldsymbol{B}$ - вектор-столбец искомых параметров (коэффициентов модели) размерности $k+1$, а $\boldsymbol{\varepsilon}$ - вектор-столбец случайных ошибок размерности n с математическим ожиданием

$$\boldsymbol{M}(\varepsilon)=0 \qquad (1.76)$$

и матрицей рассеяния

$$\boldsymbol{V}(\varepsilon) = \boldsymbol{M}(\varepsilon\varepsilon^{\mathrm{T}}) = б^2\boldsymbol{I}, \qquad (1.77)$$

где $\boldsymbol{I}$ - единичная матрица размерности $n\times n$, а $б^2$ - дисперсия ошибок наблюдений. Увеличение числа столбцов матрицы $\boldsymbol{X}$ на единицу обусловлено введением в неё первого, а, точнее, нулевого столбца, состоящего из единиц, что вызвано необходимостью вычисления свободного члена b_0. Условия (1.76) и (1.77) соответствуют предположению о том, что $\boldsymbol{\varepsilon}_1$ некоррелированы (т.е. взаимонезависимы), имеют нулевые средние и одинаковую дисперсию $б^2$. Раньше, в 50-е - 60-е годы, считалось, что на элементы марицы $\boldsymbol{X}$ не накладывается никаких ограничений. Позднее исследователи пришли к выводу, что это далеко не так. Дело в том, что независимые переменные $x_{i,j}$ в процессе экспериментов измеряются и (или) поддерживаются на требуемом уровне с помощью соответствующих приборов. А поскольку регуляторы и контрольно-измерительные приборы имеют вполне конечную погрешность, то эти ошибки могут существенно исказить окончательные результаты. Поэтому в практике экспериментальных исследований принято, наряду с возможно более высокой точностью измерений $x_{i,j}$, варьирование независимых переменных $x_{i,j}$ от опыта к опыту осуществлять с интервалами, превышающими точность измерения в 50-100 раз. Например, при экспериментальном определении влияния температуры на скорость химических реакций (или на вязкость жидкостей) с помощью лабораторного ультратермостата, поддерживающего температуру с точностью ±0.1°, эксперименты проводят с интервалом 5÷10°, иногда больше.

Метод НК состоит в минимизации скалярной суммы квадратов

$$\Phi = \sum_{i=1}^{n}\left(y_i - \sum_{j=0}^{k} b_j x_{i,j}\right)^2 = (\boldsymbol{Y}-\boldsymbol{BX})^{\mathrm{T}}(\boldsymbol{Y}-\boldsymbol{BX}) \qquad (1.78)$$

по компонентам вектора $\boldsymbol{B}$. Поскольку функция $\Phi \geqslant 0$ при любых b_0, $b_1, \ldots, b_k$, у неё должен быть хотя бы один минимум. Условием минимума является условие $\Phi / \boldsymbol{B} = 0$. Выполняя дифференцирование, получим

$$2\boldsymbol{X}^{\mathrm{T}}(\boldsymbol{Y}-\boldsymbol{XB}) = 0, \qquad (1.79)$$

откуда находим вектор-столбец оценок $\boldsymbol{B}$:

$$\boldsymbol{B} = (\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X})^{-1}(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{Y}). \qquad (1.80)$$

При этом предполагается, что матрица $(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X})$ не вырождена и, следовательно, может быть обращена. К сожалению, практически достаточно часто матрица $(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X})$ оказывается вырожденой, что, отчасти, и явилось стимулом дальнейшего развития методов поиска вектора оценок $\boldsymbol{B}$.

2. ПЛАНИРОВАНИЕ ВТОРОГО ПОРЯДКА

Если зависимость $y=f(x_1, x_2, \ldots, x_k)$ не является линейной (что является скорее правилом, чем исключением; линейных зависимостей в природе значительно меньше, чем нелинейных), а область факторного пространства достаточно велика, чтобы кривизна поверхности отклика не превышала ошибку эксперимента, то, как правило, уравнение, полученное по плану первого порядка (ПФЭ), оказывается неадекватным эксперименту [2, 9, 10, 17]. В некоторых случаях может выручить насыщение плана парными (и более) произведениями, но обычно переходят к планированию второго порядка. Как известно, через две точки можно провести только одну прямую и множество кривых линий. Очевидно, что для более или менее надёжной фиксации кривой линии необходимы как минимум три точки. Переходя к задаче получения уравнения регрессии второго порядка, можно утверждать, что для описания кривой поверхности необходимо, чтобы значения факторов $X_1, X_2, \ldots, X_k$ изменялись не менее чем на трёх уровнях. Были естественные попытки реализации планов типа 3^k, например:

$$
\begin{array}{c}
\quad z_0 \;\; z_1 \;\; z_2 \\
\mathbf{Z} = \begin{vmatrix} +1 & +1 & +1 \\ +1 & 0 & +1 \\ +1 & -1 & +1 \\ +1 & +1 & 0 \\ +1 & 0 & 0 \\ +1 & -1 & 0 \\ +1 & +1 & -1 \\ +1 & 0 & -1 \\ +1 & -1 & -1 \end{vmatrix} \Leftrightarrow \mathbf{Y} = \begin{vmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \\ y_6 \\ y_7 \\ y_8 \\ y_9 \end{vmatrix};
\end{array}
$$

Но полный факторный эксперимент типа 3^k требует большого количества опытов (см. табл. 2.1) что практически невозможно (см. также ДФЭ [17]).

Таблица 2.1.

Зависимость количества опытов и числа степеней свободы от размерности факторного пространства

k	n	ν
2	9	3
3	27	17
4	81	66
5	243	222
6	729	701
...	...	...
k	3^k	$n-l$

Д. Бокс и К. Уилсон предложили *последовательное* или *композиционное* планирование, при котором план типа 2^k после получения неадекватного уравнения (т.е. в случае неудачи) рассматривается как *ядро*, вокруг которого достраиваются специальным образом расположенные так называемые *звёздные точки* (рис. 2.1).

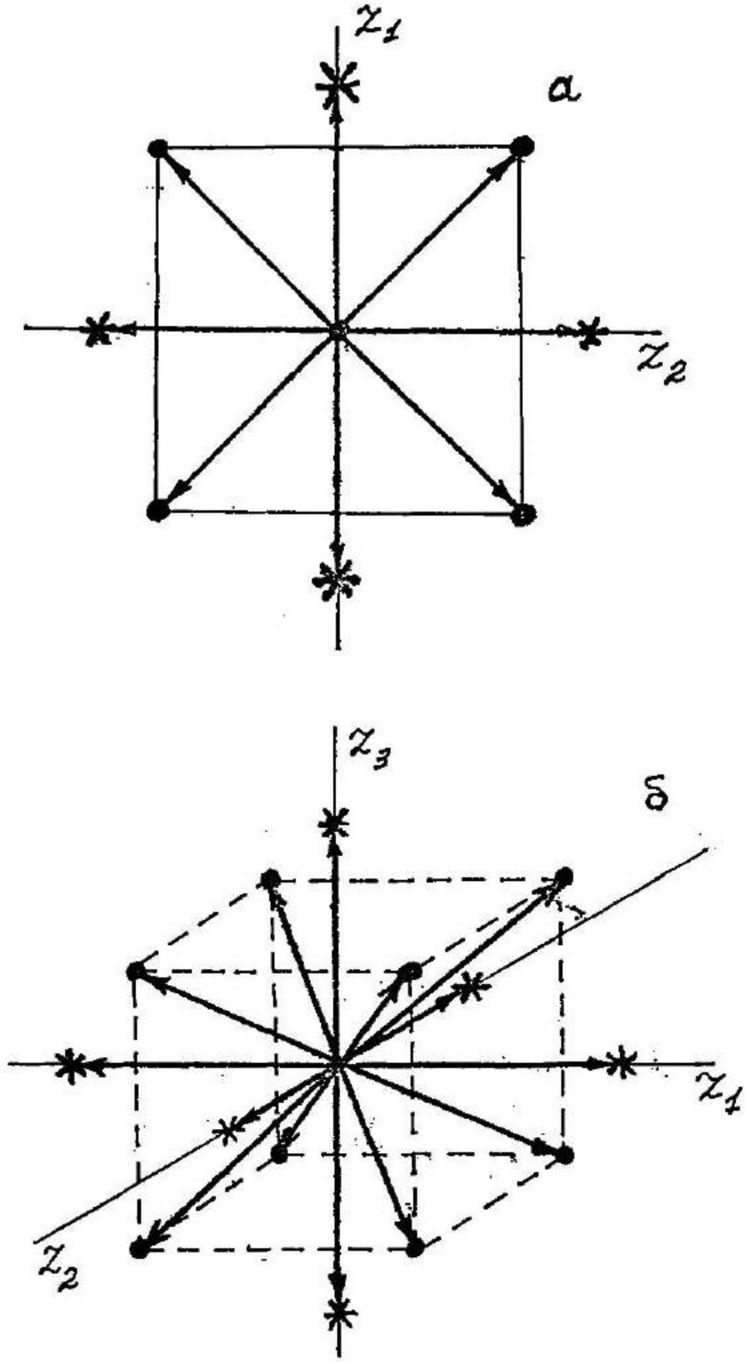


Рис. 2.1. Центральные композиционные планы экспериментов:
а – двухфакторный план; б – трёхфакторный план

Такие планы получили название *центральных композиционных планов* (ЦКП). Общее число опытов при k факторах будет равно $n=2^k+2k+n_0$, где n_0 – количество опытов в центре плана. Опыты в центре плана одновременно являются и опытами на воспроизводимость:

$$
\mathbf{Z} = \begin{array}{c} Z_0 \;\; Z_1 \;\; Z_2 \\ \left|\begin{array}{ccc}
+1 & +1 & +1 \\
+1 & -1 & +1 \\
+1 & +1 & -1 \\
+1 & -1 & -1 \\
+1 & +\alpha & 0 \\
+1 & -\alpha & 0 \\
+1 & 0 & -\alpha \\
+1 & 0 & -\alpha \\
+1 & 0 & 0 \\
\dots & \dots & \dots \\
+1 & 0 & 0
\end{array}\right| \end{array}
\Leftrightarrow \mathbf{Y} = \left|\begin{array}{c}
y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \\ y_6 \\ y_7 \\ y_8 \\ y_9 \\ \dots \\ y_n
\end{array}\right|. \qquad (2.1)
$$

Общее число опытов по центральному композиционному плану в несколько раз меньше, чем по плану типа 3^k. Главный недостаток ЦКП - невозможность совмещения *ортогональности* и *ротатабельности*. Дело в том, что выбор критерия оптимальности плана в значительной степени произволен, а величина звёздного плеча α и число опытов в центре плана n_0 непосредственно связаны с критерием оптимальности плана [2, 10]. В первых работах, посвящённых математическому описанию области близкой к экстремуму (область поверхности отклика, близкую к экстремуму, называют *почти стационарной областью*), использовалось ортогональное планирование второго порядка по аналогии с ортогональным планированием первого порядка.

2.1. Ортогональное центральное композиционное планирование

Запишем матрицу центрального композиционного плана для k=2:

$$
\mathbf{Z} = \begin{array}{c} Z_0 \;\; Z_1 \;\; Z_2 \;\; Z_1Z_2 \;\; Z_2^2 \;\; Z_2^2 \\ \left|\begin{array}{cccccc}
+1 & +1 & +1 & +1 & +1 & +1 \\
+1 & -1 & +1 & -1 & +1 & +1 \\
+1 & +1 & -1 & -1 & +1 & +1 \\
+1 & -1 & -1 & +1 & +1 & +1 \\
+1 & +\alpha & 0 & 0 & \alpha^2 & 0 \\
+1 & -\alpha & 0 & 0 & \alpha^2 & 0 \\
+1 & 0 & -\alpha & 0 & 0 & \alpha^2 \\
+1 & 0 & -\alpha & 0 & 0 & \alpha^2 \\
+1 & 0 & 0 & 0 & 0 & 0 \\
\dots & \dots & \dots & \dots & \dots & \dots \\
+1 & 0 & 0 & 0 & 0 & 0
\end{array}\right| \end{array}
\Leftrightarrow \mathbf{Y} = \left|\begin{array}{c}
y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \\ y_6 \\ y_7 \\ y_8 \\ y_9 \\ \dots \\ y_n
\end{array}\right|. \qquad (2.2)
$$

Очевидно, что матрица $\boldsymbol{Z}$ плана неортогональна:

$$\sum_{i=1}^{n} z_{i,0} z_{i,j}^{2} \neq 0; \quad j=1,\ 2, \ldots,\ k;$$

$$\sum_{i=1}^{n} z_{i,l}^{2} z_{i,m}^{2} \neq 0; \quad l \neq m; \ l,\ m=1,\ 2, \ldots,\ k, \tag{2.3}$$

так как $z_{i,0}$ всегда равно единице, а нестрицательные величины $z^2{}_{i,j}$ не могут быть все равны нулю.

Для того, чтобы матрица $\boldsymbol{Z}$ типа (2.2) стала ортогональной, очевидно, необходимо произвести некоторое преобразование столбцов $z^2{}_j$ и специальным образом выбрать величину звёздного плеча $\boldsymbol{\alpha}$.

Если столбцы $z^2{}_j$ в матрице плана преобразовать следующим образом [10]:

$$z'_{i,j} = z_{i,j}^{2} - \frac{1}{n} \cdot \sum_{i=1}^{n} z_{i,j}^{2} = z_{i,j}^{2} - \bar{z}_{j}^{2}, \tag{2.4}$$

то скалярные произведения будут равны нулю

$$\sum_{i=1}^{n} z_{i,0} z'_{i,j} = \sum_{i=1}^{n} z_{i,j}^{2} - n\bar{z}_{j}^{2} = 0, \tag{2.5}$$

а вектор-столбцы для квадратичных членов останутся пока неортогональными:

$$\sum_{i=1}^{n} z'_{i,l} z'_{i,m} \neq 0. \tag{2.6}$$

Для обеспечения полной ортогональности матрицы типа (2.2) следует произвести обращение в общем виде матрицы $\boldsymbol{Z}^{\mathrm{T}}\boldsymbol{Z}$ и определить $\boldsymbol{\alpha}$ из условия равенства нулю недиагонального элемента обратной матрицы [2]:

$$\boldsymbol{\alpha}^4 + 2^{k}\boldsymbol{\alpha}^2 - 2^{k-1}(k+0,5n_0) = 0, \ \text{ядро } 2^{k}; \tag{2.7}$$

$$\boldsymbol{\alpha}^4 + 2^{k-1}\boldsymbol{\alpha}^2 - 2^{k-2}(k+0,5n_0) = 0, \ \text{ядро } 2^{k-1}. \tag{2.8}$$

Запись условия (2.8) обусловлена тем, что при $k \geqslant 5$ уже полуреплика обеспечивает получение несмешанных оценок для линейных эффектов и эффектов парного взаимодействия (см. выше). Поскольку ортогональность принята за достаточный критерий оптимальности плана, то на количество опытов в центре плана накладывается только одно ограничение - $n_0 \geqslant 1$. Пределом n_0 является разумная достаточность достижения требуемой точности эксперимента и вычисления дисперсии воспроизводимости $s^2_{оп}$. Решения условий (2.7) и (2.8) представлены в таблице 2.1 [2]:

Таблица 2.1.

Количества экспериментов в центре плана n_0
и соответствующие им величины звёздных плеч α

n_0	Количество факторов k							
	2		3		4		5*	
	α^2	α	α^2	α	α^2	α	α^2	α
1	1.000	1.000	1.476	1.215	2.000	1.414	2.390	1.546
2	1.160	1.077	1.650	1.285	2.164	1.471	2.580	1.606
3	1.317	1.148	1.831	1.353	2.390	1.546	2.770	1.664
4	1.475	1.214	2.000	1.414	2.580	1.606	2.950	1.718
5	1.606	1.267	2.164	1.471	2.770	1.664	3.140	1.772
6	1.742	1.320	2.325	1.525	2.950	1.718	3.310	1.819
7	1.873	1.369	2.481	1.575	3.140	1.772	3.490	1.868
8	2.000	1.414	2.633	1.623	3.310	1.819	3.660	1.913
9	2.113	1.454	2.782	1.668	3.490	1.868	3.830	1.957
10	2.243	1.498	2.928	1.711	3.660	1.913	4.000	2.000

С учётом вышеизложенного матрица центрального композиционного плана для k=2 и для n_0=1 будет иметь вид

$$
\begin{array}{c}
\begin{array}{cccccc} z_0 & z_1 & z_2 & z_1 z_2 & z_2' & z_2' \end{array} \\
\boldsymbol{Z} = \left|\begin{array}{cccccc}
+1 & +1 & +1 & +1 & +1/3 & +1/3 \\
+1 & -1 & +1 & -1 & +1/3 & +1/3 \\
+1 & +1 & -1 & -1 & +1/3 & +1/3 \\
+1 & -1 & -1 & +1 & +1/3 & +1/3 \\
+1 & +\alpha & 0 & 0 & +1/3 & -2/3 \\
+1 & -\alpha & 0 & 0 & +1/3 & -2/3 \\
+1 & 0 & -\alpha & 0 & -2/3 & +1/3 \\
+1 & 0 & -\alpha & 0 & -2/3 & +1/3 \\
-1 & 0 & 0 & 0 & -2/3 & -2/3
\end{array}\right| \Leftrightarrow \boldsymbol{Y} = \left|\begin{array}{c} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \\ y_6 \\ y_7 \\ y_8 \\ y_9 \end{array}\right|.
\end{array}
\qquad (2.9)
$$

Уравнение регрессии при ортогональном центральном композиционном планировании получают в виде [2, 10]

$$
\hat{y} = b_0' + \sum_{j=1}^{k} b_j z_j + \sum_{\substack{l=1 \\ m=2 \\ l \neq m}}^{k} b_{l,m} z_l z_m + \sum_{j=1}^{k} b_{jj} z_j'; \qquad (2.10)
$$

где

$$z_j^{'}=z_j^2 - \frac{1}{n}\cdot\sum_{i=1}^{n} z_{i.j}^2. \qquad (2.11)$$

Вследствие ортогональности плана все коэффициенты уравнения (2.10) определяются независимо друг от друга по формулам

$$b_0^{'} = \frac{1}{n}\cdot\sum_{i=1}^{n} y_i; \qquad (2.12)$$

$$b_j=\sum_{i=1}^{n} z_{i.j}y_i \Big/ \sum_{i=1}^{n} z_{i.j}^2, \quad (j\neq 0); \qquad (2.13)$$

$$b_{l.m} = \sum_{i=1}^{n} z_{i.l}z_{i.m}y_i \Big/ \sum_{i=1}^{n} z_{i.l}^2 z_{i.m}^2, \quad l\neq m,\ l\neq 0,\ m\neq 0; \qquad (2.14)$$

$$b_{j.j}=\sum_{i=1}^{n} z_{i.j}^{'}y_i \Big/ \sum_{i=1}^{n}\left(z_{i.j}^{'}\right)^2. \qquad (2.15)$$

Подставляя в уравнение (2.10) выражение (2.11) для преобразованной переменной $z^{'}_j$, получим привычное выражение для уравнения регрессии

$$\hat{y} = b_0 + \sum_{j=1}^{k} b_j z_j + \sum_{\substack{l=1\\ m=2}}^{k} b_{l.m} z_l z_m + \sum_{j=1}^{k} b_{jj} z_j^2, \quad l\neq m,\ l\neq 0,\ m\neq 0, \qquad (2.16)$$

где

$$b_0= b_0^{'} - \frac{1}{n}\cdot\sum_{j=1}^{k} b_{jj} \cdot\sum_{i=1}^{n} z_{i.j}^2. \qquad (2.17)$$

2.1.1. Проверка коэффициентов уравнения на значимость

Проверка коэффициентов уравнения на значимость по существу сводится к ответу на вопрос, влияет j-ный фактор на процесс или не влияет. Точно ответить на этот вопрос, как правило, не удаётся, поэтому рассматриваются две гипотезы: коэффициент $b_j=0$ $(H_0: b_j=0)$ и альтернативная, $b_j\neq 0$ $(H_1: b_j\neq 0)$, а ответ на вопрос о значимости формулируется с той или иной *доверительной вероятностью*. Независимо от абсолютного значения коэффициента незначимое отличие от нуля будет соответствовать факту нахождения нуля в *доверительном интервале* коэффициента b_j

$$b_j - s_{b_j} \cdot t^{\alpha}_{\nu_{оп}} \leqslant \beta_j \leqslant b_j + s_{b_j} \cdot t^{\alpha}_{\nu_{оп}}, \qquad (2.18)$$

где β_j - генеральный j-тый коэффициент; s_{bj} - *стандартное отклонение* коэффициента b_j; t^{α} - *критерий Стьюдента* для *уровня значимости* α и *числа степеней свободы* $\nu_{оп}$ (сравните с соотношением 1.53). Практически *нулевая гипотеза* формулируется как отсутствие различия между коэффициентом b_j и его *стандартным отклонением*, а соответствующий критерий конструируется аналогично (1.45) и (1.54):

$$t_{оп} = \frac{|b_j|}{s_{b_j}} \Leftrightarrow t^{\alpha}_{\nu_{оп}}, \qquad (2.19)$$

где t - табличный *критерий Стьюдента* для *числа степеней свободы* ν дисперсии воспроизводимости и *уровня значимости* α (подразумевается, что в формулу подставляется положительное значение стандартного отклонения s_{bj}). Поскольку коэффициент b_j может быть как положительным, так и отрицательным, следует использовать табличное значение двустороннего критерия Стьюдента.

Поскольку диагональные элементы информационной матрицы ($\boldsymbol{Z}^{\mathrm{T}}\boldsymbol{Z}$) неодинаковы, дисперсии всех коэффициенты определяются с разной точностью по формулам [2, 10, 11]

$$s^2_{b_0} = s^2_{b_0'} + s^2_{b_{jj}} \cdot \frac{k}{n}\,(2^k + 2\alpha^2), \text{ где } s^2_{b_0'} = \frac{1}{n}\, s^2_{оп}; \qquad (2.20)$$

$$s^2_{b_j} = s^2_{оп} \Big/ \sum_{i=1}^{n} z^2_{i,j}, \quad \text{где} \sum_{i=1}^{n} z^2_{i,j} = 2^k + 2\alpha^2, \quad j \neq 0; \qquad (2.21)$$

$$s^2_{b_{l,m}} = s^2_{оп} \Big/ \sum_{i=1}^{n} z^2_{i,l} z^2_{i,m}, \quad \text{где} \sum_{i=1}^{n} z^2_{i,l} z^2_{i,m} = 2^k, \quad l \neq m; \qquad (2.22)$$

$$s^2_{b_{j,j}} = s^2_{оп} \Big/ \sum_{i=1}^{n} \left(z'_{i,j}\right)^2. \qquad (2.23)$$

При определении дисперсии воспроизводимости по опытам в центре плана расчёты следует выполнять по формулам (1.16) и (1.17).

Если опытный критерий Стьюдента меньше табличного, то с доверительной вероятностью $P=1-\alpha$ нулевая гипотеза принимается, т.е. коэффициент b_j незначимо отличается от нуля. Если опытный критерий Стьюдента больше табличного, он попадает в критическую область, в область отклонения нулевой гипотезы (см. рис.1.3), то следует принять альтернативную гипотезу $H_1: b_j \neq 0$, т.е. коэффициент b_j значимо отличается от нуля и j-тый фактор влияет на процесс. Вероятность риска от принятия такого решения не больше принятого уровня значимости α.

2.1.2. Проверка уравнения регрессии на адекватность

Проверка уравнения регрессии на адекватность производится после отсева незначимых коэффициентов и сводится к проверке однородности дисперсии воспроизводимости и дисперсии адекватности. Нулевая гипотеза формулируется как гипотеза об отсутствии различия между этими дисперсиями:

$$F^{\text{оп}}_{\nu_{\text{ад}},\nu_{\text{оп}}} = \frac{S^2_{\text{ад}}}{S^2_{\text{оп}}} \Leftrightarrow F^{\alpha}_{\nu_{\text{ад}},\nu_{\text{оп}}}, \tag{2.24}$$

где $\nu_{\text{ад}}$ - число степеней свободы дисперсии адекватности, равное $\nu=n-l$; l - число связей, наложенных на выборку (общее число значимых коэффициентов в уравнении регрессии); $\nu_{\text{оп}}$ - число степеней свободы дисперсии воспроизводимости. Опытный критерий Фишера сравнивается с табличным значением для принятого уровня значимости α. Если $F^{\text{оп}}<F^{\alpha}$, то оснований для отклонения нулевой гипотезы об отсутствии различия дисперсий $s^2_{\text{ад}}$ и $s^2_{\text{оп}}$ нет, т.е. дисперсии однородны и уравнение регрессии адекватно. Если $F^{\text{оп}}>F^{\alpha}$, то нулевую гипотезу следует отклонить, т.е. дисперсии неоднородны и уравнение регрессии неадекватно. Вероятность риска от принятия такого решения не больше принятого уровня значимости α.

Дисперсия адекватности $s^2_{\text{ад}}$ вычисляется по следующей формуле:

$$s^2_{\text{ад}} = \frac{1}{\nu_{\text{ад}}} \cdot \sum_{i=1}^{n} (y_i - \hat{y}_i)^2, \tag{2.25}$$

где $\hat{y}$ значение функции отклика, рассчитываемое по уравнению (2.16).

2.1.3. Декодирование уравнения регрессии

Декодирование уравнения регрессии производится с помощью соотношения используемого при кодировании переменных:

$$z_j = \frac{x_j - x_j^0}{\Delta x_j}; \qquad (2.26)$$

аналогично ПФЭ. Также следует напомнить, что бывают случаи незначимости коэффициента b_j и значимости, например, b_{jj}. Согласно требованиям регрессионного анализа незначимый коэффициент b_j необходимо исключить, а значимый коэффициент b_{jj} оставить, но при декодировании уравнения коэффициент при j-том факторе появится и j-тый фактор как бы возвращается в уравнении регрессии. Вопрос с действительной значимостью j-того фактора остаётся открытым.

2.1.4. Пример ортогонального планирования

Для исследования влияния соляной кислоты, формалина, наполнителя и температуры на время затвердевания гипана (гипан - гидролизованный полиакрилонитрил) произведено ортогональное центральное композиционное планирование. При приготовлении тампонажного материала используются раствор соляной кислоты, формалин и наполнитель в соответствующих весовых частях на 100 весовых частей гипана.

Обозначения и пределы изменения переменных:
$x_1 = c_{HCl}$, 37%-ный водный раствор HCl, 10÷50 в.ч./100 в.ч. гипана;
$x_2 = c_ф$, формалин, 50÷5 в.ч./100 в.ч. гипана;
$x_3 = c_н$, наполнитель, 0÷25 в.ч./100 в.ч. гипана;
$x_4 = T$, температура, T=293÷343 K.

При отладке методики эксперимента достигнута точность эксперимента, соответствующая дисперсии воспроизводимости $s^2_{оп}=0,56$ с числом степеней свободы $\nu=3$.

Требуется получить уравнение регрессии вида:

$$y = b_0 + b_1x_1 + b_2x_2 + b_3x_3 + b_4x_4 + b_{1,2}x_1x_2 + b_{1,3}x_1x_3 + b_{1,4}x_1x_4 +$$
$$+ b_{2,3}x_2x_3 + b_{2,4}x_2x_4 + b_{3,4}x_3x_4 + b_{11}x^2{}_1 + b_{22}x^2{}_2 + b_{33}x^2{}_3 + b_{44}x^2{}_4.$$

Кодирование переменных

Координаты центра плана и интервал варьирования переменных вычисляются по формулам

$$x_j^0 = \frac{x_j^{max} + x_j^{min}}{2}; \qquad \Delta x_j = \frac{x_j^{max} - x_j^{min}}{2}.$$

Величину звёздного плеча α, соответствующую числу факторов $k=4$ и количеству предварительно поставленных экспериментов в центре плана n_0, находим по табл. 2.1. Результаты кодирования представлены в табл. 2.2.

Таблица 2.2

Результаты кодирования факторов и значения независимых переменных в исходной и кодированной системах координат

Компоненты ортогонального центрального композиционного плана	Значения факторов							
	Содержание 37% HCl		Содержание формалина		Содержание наполнителя		Температура	
	Физические	Код. z_1	Физические	Код. z_2	Физические	Код. z_3	Физические	Код. z_4
Центр плана	30	0	27.5	0	12.5	0	318	0
Интервал варьирования	12,4	1	14,0	1	7,8	1	15	1
Звёздное плечо	20	1,607	22,5	1,607	12,5	1,607	25	1,607
Нижний уровень	17,6	-1	13,5	-1	4,7	-1	303	-1
Верхний уровень	42,4	+1	41,5	+1	20,3	+1	334	+1

Матрица четырёхфакторного плана, результаты экспериментов и результаты расчётов по искомому уравнению (см. ниже) представлены в табл. 2.3.

Вычисление коэффициентов b_0, b_j, $b_{l,m}$ и b_{jj} производится по формулам (2.12), (2.13), (2.14), (2.15) и (2.17), результаты представлены в табл. 2.4.

Таблица 2.3

Матрица плана, результаты экспериментов и расчётов по искомому уравнению регрессии

Z_0	Z_1	Z_2	Z_3	Z_4	$y_{i,\text{эксп}}$	$y_{i,\text{расч}}$
+1	+1	+1	+1	+1	7,25	8,7555
+1	−1	+1	+1	+1	1,5	1,1616
+1	+1	−1	+1	+1	13,7	12,5735
+1	−1	−1	+1	+1	7,5	9,2546
+1	+1	+1	−1	+1	5,0	5,8467
+1	−1	+1	−1	+1	2,0	1,1278
+1	+1	−1	−1	+1	4,5	5,7897
+1	−1	−1	−1	+1	4,0	5,3458
+1	+1	+1	−1	−1	7,25	6,1843
+1	−1	+1	+1	−1	42,7	42,44
+1	+1	−1	+1	−1	3,0	4,9023
+1	−1	−1	+1	−1	46,0	45,4334
+1	+1	+1	−1	−1	8,0	7,2755
+1	−1	+1	−1	−1	45,0	46,4067
+1	+1	−1	−1	−1	1,5	2,1185
+1	−1	−1	−1	−1	46,0	45,5247
+1	+1.607	0	0	0	2,0	0,5964
+1	−1.607	0	0	0	30,0	29,3746
+1	0	+1.607	0	0	10,0	10,3058
+1	0	−1.607	0	0	15,0	12,6652
+1	0	0	+1.607	0	5,75	6,1046
+1	0	0	−1.607	0	4,5	3,8423
+1	0	0	0	+1.607	3,0	0,875
+1	0	0	0	−1.607	31,0	31,096
+1	0	0	0	0	5,25	4,9744
+1	0	0	0	0	4,2	4,9744
+1	0	0	0	0	3,9	4,9744
+1	0	0	0	0	5,4	4,9744

После вычисления коэффициентов при квадратичных членах необходимо произвести пересчёт коэффициента b_0 по формуле (2.17):

$$b_0 = b_0' - \frac{1}{n} \cdot \sum_{j=1}^{k} b_{jj} \cdot \sum_{i=1}^{n} z_{i,j}^2 = 4,9744.$$

Таблица 2.4

Коэффициенты уравнения с кодированными переменными, значения экспериментального критерия Стьюдента и коэффициенты уравнения с физическими переменчыми

Коэф. b_j	Коэффициенты кодированного уравнения	Эксп. критерий Стьюдента	Коэффициенты декодированного уравнения
b'_0	13,03	92,1	2722,6
b_1	-8,9531	55,0	-21,24
b_2	-0,734	4,51	1,097
b_3	0,7044	4,33	-2,607
b_4	-9,402	57,8	-14,385
$b_{1.2}$	1,069	5,71	0,006156
$b_{1.3}$	0,719	3,84	0,007431
$b_{1.4}$	10,96	54,6	0,05894
$b_{2.3}$	-0,9688	5,17	-0,008871
$b_{2.4}$	-1,275	6,81	-0,006071
$b_{3.4}$	1,0000	5,34	0,008547
$b_{1.1}$	3,876	18,9	0,0252
$b_{2.2}$	2,521	12,3	0,01286
$b_{3.3}$	-0,3344	1,63	-
$b_{4.4}$	4,263	20,8	0,01895

Дисперсия воспроизводимости определяется по формулам (1.16) и (1.17):

$$\bar{y}_0 = \frac{1}{n_0} \cdot \sum_{i=1}^{n_0} y_{0.i} = \frac{18,75}{4} = 4,69;$$

$$s^2_{оп} = \frac{1}{(n_0-1)} \sum_{i=1}^{n_0} (y_{0.i} - \bar{y}_0)^2 = \frac{1,6819}{3} = 0,56.$$

Дисперсия коэффициента b_0 определяются по формулам (2.20):

$$s^2_{b'_0} = 0,02002; \qquad s^2_{b_0} = 0,1153.$$

Дисперсии остальных коэффициентов уравнения регрессии определяются по формулам (2.21), (2.22) и (2.23):

$$s^2_{b_j} = 0,02649; \qquad s^2_{b_{l,m}} = 0,03504; \qquad s^2_{b_{j,j}} = 0,042014.$$

Соответствующие квадратичные отклонения коэффициентов уравнения регрессии представлены ниже:

$$s_{b'_0} = \pm 0,1415; \qquad s_{b_0} = \pm 0,3396; \qquad s_{b_j} = \pm 0,1627;$$

$$s_{b_{l,m}} = \pm 0,1872; \qquad s_{b_{j,j}} = \pm 0,205.$$

Вычисленные по формуле (2.19) опытные критерии Стьюдента для каждого коэффициента уравнения регрессии представлены в табл. 2.4. Для уровня значимости α=0,05 и числа степеней свободы дисперсии воспроизводимости ν=3 табличное значение критерия Стьюдента t=3,18 (Табл. П1). Результаты сравнения опытных критериев Стьюдента с табличным значением показывают, что, кроме t_{33}, все опытные значения t-критериев попадают в критическую область (см. рис. 1.3), область отклонения нулевой гипотезы об отсутствии различия коэффициента b_j и его стандартного отклонения s_{bj}. Опытный t-критерий коэффициента b_{33} меньше критического табличного значения 3,18. В этом случае у нас нет оснований отклонить нулевую гипотезу об отсутствии различия между коэффициентом b_{33} и его стандартным отклонением s_{b33}, т.е. коэффициент b_{33} незначим и его следует исключить из уравнения регрессии. Таким образом, уравнение регрессии с кодированными переменными имеет вид:

$$y = b_0 + b_1 z_1 + b_2 z_2 + b_3 z_3 + b_4 z_4 + b_{1,2} z_1 z_2 + b_{1,3} z_1 z_3 + b_{1,4} z_1 z_4 +$$

$$+ b_{2,3} z_2 z_3 + b_{2,4} z_2 z_4 + b_{3,4} z_3 z_4 + b_{11} z^2{}_1 + b_{22} z^2{}_2 + b_{44} z^2{}_4.$$

Проверка уравнения регрессии на адекватность осуществляется по соотношению (2.24). Предварительно вычислим дисперсию адекватности по уравнению (2.25):

$$s^2_{ад} = \frac{1}{\nu_{ад}} \cdot \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = \frac{25,491}{14} = 2,496.$$

Опытный критерий Фишера

$$F^{оп}_{\nu_{ад}, \nu_{оп}} = \frac{s^2_{ад}}{s^2_{оп}} = \frac{2,496}{0,5606} = 4,45,$$

где $\nu_{ад}$=14, $\nu_{оп}$=3.

Опытный критерий Фишера сравниваем с табличным (Табл.П2):

$$F^{0,05}_{14,3} = 8,7 > F^{оп}_{14,3} = 4,45.$$

Опытный критерий Фишера не попадает в критическую область (см. рис. 1.2), он меньше табличного значения, взятого для уровня значимости α=0,05. Это значит, что с доверительной вероятностью P=0,95 нулевая гипотеза об однородности дисперсий $s^2_{ад}$ и $s^2_{оп}$ не противоречит результатам проверки, т.е. уравнение регрессии адекватно.

Декодирование уравнения регрессии

Декодирование уравнения регрессии производится с помощью соотношения (2.26), используемого при кодировании переменных:

$$y = b_0 + b_1\left(\frac{x_1 - x_1^0}{\Delta x_1}\right) + b_2\left(\frac{x_2 - x_2^0}{\Delta x_2}\right) + b_3\left(\frac{x_3 - x_3^0}{\Delta x_3}\right) + b_4\left(\frac{x_4 - x_4^0}{\Delta x_4}\right) +$$

$$+ b_{1,2}\left(\frac{x_1 - x_1^0}{\Delta x_1}\right)\left(\frac{x_2 - x_2^0}{\Delta x_2}\right) + b_{1,3}\left(\frac{x_1 - x_1^0}{\Delta x_1}\right)\left(\frac{x_3 - x_3^0}{\Delta x_3}\right) + b_{1,4}\left(\frac{x_1 - x_1^0}{\Delta x_1}\right)\left(\frac{x_4 - x_4^0}{\Delta x_4}\right) +$$

$$+ b_{2,3}\left(\frac{x_2 - x_2^0}{\Delta x_2}\right)\left(\frac{x_3 - x_3^0}{\Delta x_3}\right) + b_{2,4}\left(\frac{x_2 - x_2^0}{\Delta x_2}\right)\left(\frac{x_4 - x_4^0}{\Delta x_3}\right) + b_{3,4}\left(\frac{x_3 - x_3^0}{\Delta x_3}\right)\left(\frac{x_4 - x_4^0}{\Delta x_4}\right) +$$

$$+ b_{11}\left(\frac{x_1 - x_1^0}{\Delta x_1}\right)^2 + b_{22}\left(\frac{x_2 - x_2^0}{\Delta x_2}\right)^2 + b_{44}\left(\frac{x_4 - x_4^0}{\Delta x_4}\right)^2.$$

Раскрывая скобки, приводя подобные члены и подставляя численные значения коэффициентов уравнения, координат центра плана и интервалов варьирования факторов, получим

$$y=2723-21,24x_1+1,1x_2-2,61x_3-14,4x_4+0,00616x_1x_2+$$
$$+0,00743x_1x_3+0,059x_1x_4-0,00887x_2x_3-0,00607x_2x_4+$$
$$+0,00855x_3x_4+0,0252x_1^2+0,013x_2^2+0,019x_4^2.$$

Полученное уравнение описывает зависимость времени затвердевания гипана от температуры и количеств соляной кислоты, формалина и наполнителя (в.ч. на 100 в.ч. гипана).

2.2. Ротатабельные планы второго порядка Бокса-Хантера

Практическое использование ортогональных планов второго порядка показало их невысокую эффективность при описании сложных поверхностей отклика. Критерий ортогональности, как критерий оптимальности плана, приводит к тому, что коэффициенты уравнения регрессии определяются с разной степенью точности. С другой стороны, при описании поверхности отклика, близкой к экстремуму, исследователя больше интересует оценка уравнения в целом, нежели оценки отдельных коэффициентов регрессии или вклад каждого из них. Это значит, что критерий оптимальности для планирования эксперимента должен быть принят на основе близости оценок к действительной поверхности отклика в возможно большей области факторного пространства. В этой связи идея рассмотрения информационных контуров (кривых или поверхностей равной информации) оказалась плодотворной (рис.2.2). Информационные контуры вычислялись с помощью выражения для дисперсии, получаемого на основе закона сложения ошибок [10]. Очевидно, что факторное пространство по своей информативности очень неоднородно: контуры не образуют концентрических колец и различные направления факторного пространства формируют разный информационный потенциал. Отношение информационных потенциалов в направлениях грани, ребра и угла плана достигает двух - пяти раз. С точки зрения экспериментатора такая информационная асимметрия крайне нежелательна, поскольку он заранее не знает того направления факторного пространства, которое в дальнейшем будет представлять преимущественный интерес.

Рассмотрение этих информационных контуров объясняет некоторые парадоксы, возникающие в практике математической обработки экспериментальных результатов. В пространстве, образованном независимыми

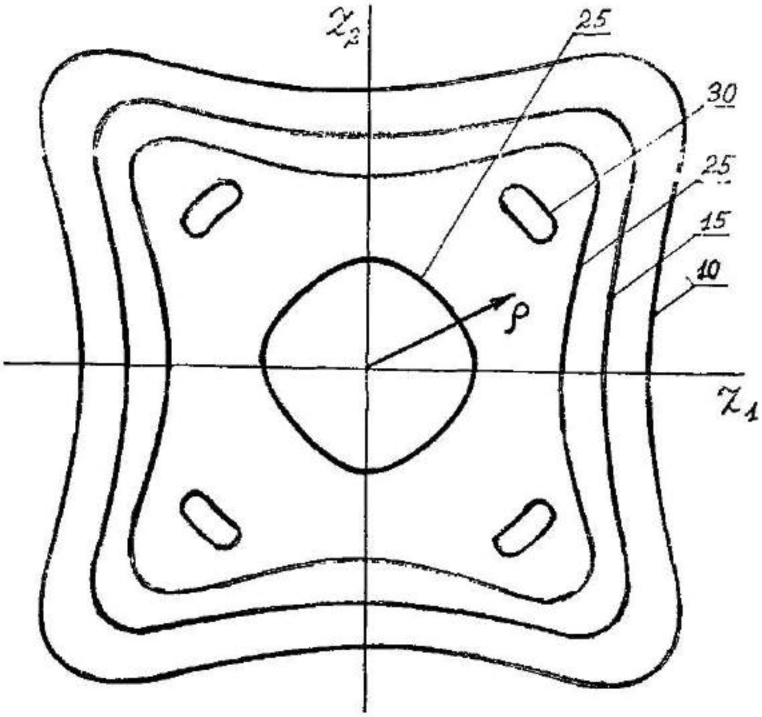


Рис. 2.2. Информационные контуры (кривые равной информации) для двухфакторного планирования

переменными и функцией отклика, бывают "провалы", в которых не удаётся получить удовлетворительного математического описания зависимости $y=f(z_1, z_2)$. Если экспериментальные "точки" попадают на места провалов, то происходят странные на первый взгляд вещи: вырождается матрица ($\mathbf{X}^T\mathbf{X}$), возрастает дисперсия и (или) корреляция отдельных коэффициентов, резко увеличивается дисперсия адекватности и т.п., т.е. резко падает точность математического описания в этих участках. Правильнее сказать, что это не есть особенность факторного пространства, а, скорее, особенность математических методов и специфика численных методов обработки данных. При этом иногда достаточно сместить нулевую точку (от °С перейти к абсолютной шкале или наоборот и т.п.), изменить систему отсчёта, например изменить масштаб переменных (перейти от *атм* к Па, от *час* к *с* или наоборот), и информационные провалы выравниваются, заполняются поверхностью отклика. Другими словами, удаётся получить адекватное математическое описание. Последнее обстоятельство объясняется тем, что факторное пространство неинвариантно к масштабным преобразованиям системы координат.

В 1957 г. G.E.P. Box и J.S. Hunter предложили считать оптимальным планированием второго порядка ротатабельное планирование, позволяющее получать симметричные информационные контуры [15]. Такой критерий хорошо согласуется с интуитивным представлением исследователя о том, чтобы информация, получаемая с помощью уравнения регрессии, была равномерно распределена по сферам факторного пространства. Ротатабельным будет такое планирование, у которого ковариационная матрица $(\mathbf{Z}^T\mathbf{Z})^{-1}$ инвариантна к ортогональному вращению координат. Это условие удовлетворяется, если моменты информационной матрицы удовлетворяют соотношениям

$$\sum_{i=1}^{n} z_{i,j}^{2} = n\lambda_2, \quad j=1,\ 2,\ldots,\ k; \tag{2.27}$$

$$\sum_{i=1}^{n} z_{i,j}^{4} = 3\cdot\sum_{i=1}^{n} z_{i,l}^{2} z_{i,m}^{2} = 3n\lambda_4; \quad l,\ m=1,\ 2,\ldots,\ k;\ l\neq m, \tag{2.28}$$

где λ_2 и λ_4 - произвольно выбираемые константы, удовлетворяющие неравенству

$$\frac{\lambda_4}{\lambda_2^2} > \frac{k}{k+2}, \tag{2.29}$$

которое является условием невырожденности матрицы $\mathbf{Z}^T\mathbf{Z}$. Остальные моменты информационной матрицы равны нулю. В зависимости от значений λ_2 и λ_4 изменяется вид информационных профилей. Так, например, информационный потенциал выборки возрастает с увеличением λ_4 и уменьшается с увеличением радиуса сферы ρ, на которой находятся точки ядра плана (рис. 2.3). Более того, при повороте системы координат в факторном пространстве дисперсии коэффициентов изменяются, в том числе дисперсии коэффициентов $b_{l,m}$ и b_{jj} изменяются по разному (рис.2.4, а, б). Коэффициенты корреляции эффектов парного взаимодействия $b_{l,u}$ и $b_{l,m}$ и квадратичных эффектов b_{ll} и b_{mm} также изменяются при повороте системы координат в факторном пространстве от 0° до 90° (рис.2.5, а, б). Было предложено значения λ_4 выбирать так, чтобы информационный потенциал был приблизительно постоянным в объёме факторного пространства радиусом от 0 до 1, т.е. $0<\rho<1$. Такое планирование получило название *униформ-ротатабельного планирования*. Построение униформ-ротатабельного плана - сложная математическая задача [2, 10], выходящая за рамки настоящего пособия, поэтому рассмотрим выводы и практические расчётные формулы.

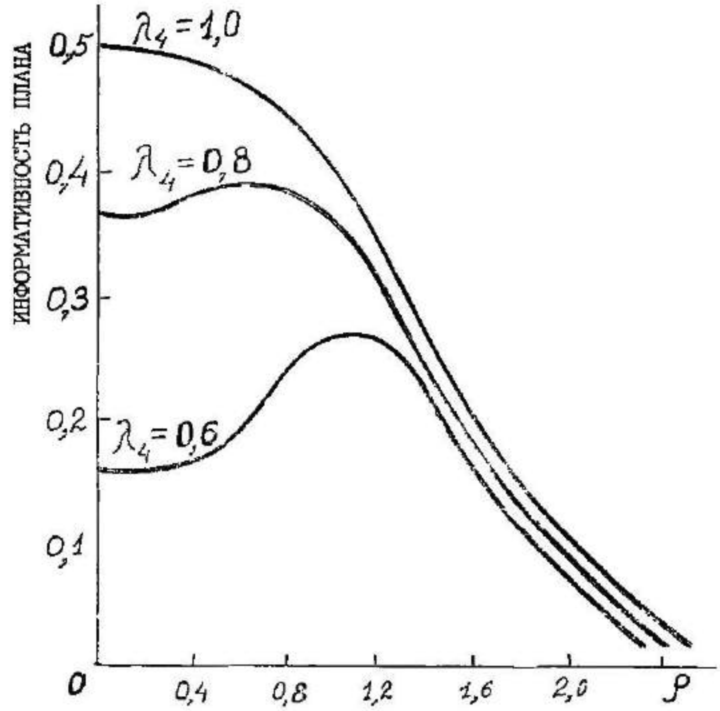


Рис. 2.3. Зависимость количества информации (относительной точности описания функции отклика) от радиуса ядра плана ρ при разных значениях λ_4 для двухфакторного ротатабельного планирования

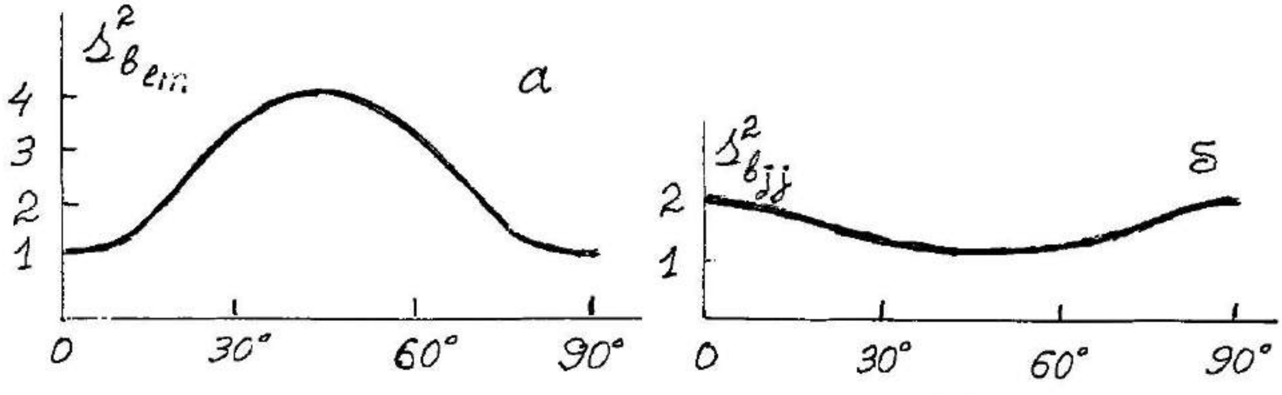


Рис. 2.4. Изменения дисперсии - при парных произведениях факторов (а), при квадратичных членах (б) коэффициентов уравнения регрессии при вращении системы координат

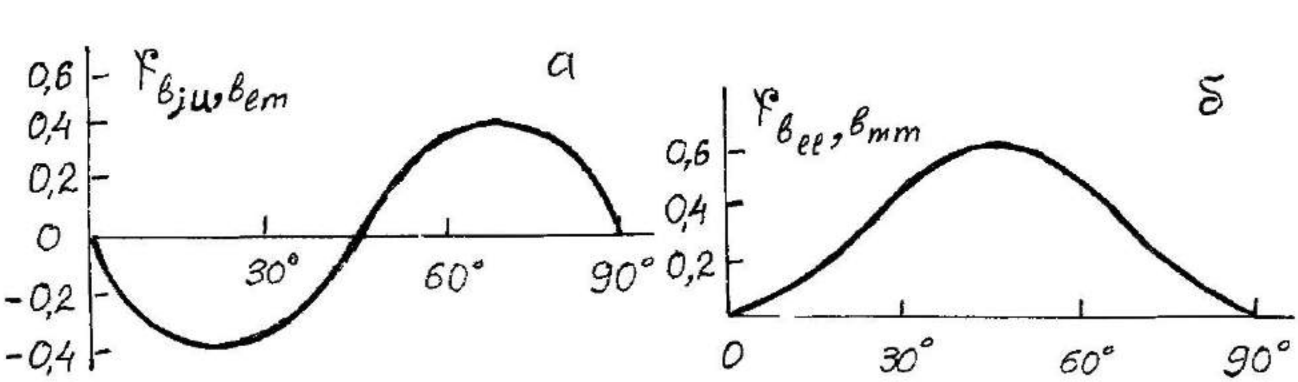


Рис. 2.5. Изменения коэффициента корреляции - при парных произведениях факторов (а), при квадратичных членах (б) коэффициентов уравнения регрессии при вращении системы координат

Определитель матрицы $\boldsymbol{Z}^{\mathrm{T}}\boldsymbol{Z}$ равен нулю в том случае, если

$$\frac{\lambda_4}{\lambda_2^2} = \frac{k}{k+2}. \qquad (2.30)$$

Рассматривая радиус сферы ρ_i на которой расположена i-тая точка в кодированном k-мерном пространстве

$$\lambda_2 = \frac{\sum_{i=1}^{n}\rho_i^2}{kn}; \qquad \lambda_4 = \frac{\sum_{i=1}^{n}\rho_i^4}{kn(k+2)}, \qquad (2.31)$$

прийдём к выводу, что условие (2.30) выполняется, если все n точек ротатабельного плана расположены на одной сфере. Поэтому для успешного решения задачи необходимо, чтобы точки ядра плана и звёздные точки были расположены на разных сферах. Другими словами, недопустимо равенство радиуса сферы ядра плана $\rho_я$ и радиуса сферы, на которой расположены звёздные точки, ρ_α, т.е. $\rho_я \neq \rho_\alpha$. Кроме этого, чтобы информационная матрица была невырожденной, в план эксперимента вводят точки, лежащие на сфере с нулевым радиусом $\rho_0=0$, т.е. n_0 точек в центре план. Общее число опытов $n=n_я+n_\alpha+n_0$, где число опытов в ядре плана $n_я=2^k$, число опытов в звёздных точках $n_\alpha=2k$, а число опытов в центре плана n_0 определяется исходя из условия обеспечения возможно большей ортогональности плана. Так, при $\lambda_4=1$ ротатабельный план оказывается почти ортогональным в том смысле, что не равна нулю только одна ковариация:

$$\mathrm{cov}_{b_0, b_{jj}} \neq 0, \qquad (2.32)$$

т.е. свободный член коррелирован только с квадратичными эффектами (дело в том, что при ортогональном планировании корреляционная матрица $(X^{\mathrm{T}}X)^{-1}$ диагональна - члены главной диагонали характеризуют дисперсии коэффициентов b_j, а все остальные члены равны нулю. В результате все коэффициенты регрессии определяются независимо друг от друга и все ковариации оказываются равными нулю. Ковариации являются количественной мерой той неопределённости, которая возникает в силу того, что коэффициенты уравнения регрессии определяются не независимо друг от друга). Выполнение условия постоянства информационного потенциала плана в возможно большем объёме факторного пространства, $0 \leqslant \rho \leqslant 1$, требует, чтобы значение λ_4 было немного меньше единицы. Это приводит к дальнейшей потере ортогональности – отличной от нуля становится ещё и ковариация

$$\mathrm{cov}_{b_{ll}, b_{mm}} \neq 0, \tag{2.33}$$

т.е. квадратичные эффекты становятся коррелированными не только со свободным членом, но и между собой. Возможно более полное выполнение всех вышеперечисленных требований приводит к тому, что в отличие от ортогонального планирования в ротатабельных униформ-планах нет свободы выбора величины звёздного плеча α и числа опытов в центре плана n_0. Величина звёздного плеча α может быть определена с помощью соотношения (2.28):

при $k<5$
$$2^k+2\alpha^4 = 3\cdot 2^k \rightarrow \alpha = 2^{k/4}; \tag{2.34}$$
при $k \geqslant 5$
$$2^{k-1}+2\alpha^4 = 3\cdot 2^{k-1} \rightarrow \alpha = 2^{(k-1)/4}. \tag{2.35}$$

В табл. 2.5 приведены значения радиуса сферы, на которой расположены точки ядра плана $\rho_я$, величины звёздного плеча α и количества опытов в центре плана n_0 [2].

Например, матрица трёхфакторного ротатабельного плана имеет вид

$$\begin{array}{cc}
 & \begin{array}{cccc} z_0 & z_1 & z_2 & z_3 \end{array} \\
\boldsymbol{Z} = & \left|\begin{array}{cccc}
+1 & +1 & +1 & +1 \\
+1 & -1 & +1 & +1 \\
+1 & +1 & -1 & +1 \\
+1 & -1 & -1 & +1 \\
+1 & +1 & +1 & -1 \\
+1 & -1 & +1 & -1 \\
+1 & +1 & -1 & -1 \\
+1 & -1 & -1 & -1 \\
+1 & +1.68 & 0 & 0 \\
+1 & -1.68 & 0 & 0 \\
+1 & 0 & +1.68 & 0 \\
+1 & 0 & -1.68 & 0 \\
+1 & 0 & 0 & +1.68 \\
+1 & 0 & 0 & -1.68 \\
+1 & 0 & 0 & 0 \\
+1 & 0 & 0 & 0 \\
+1 & 0 & 0 & 0 \\
+1 & 0 & 0 & 0 \\
+1 & 0 & 0 & 0 \\
+1 & 0 & 0 & 0
\end{array}\right|
\end{array}
\Leftrightarrow
\begin{array}{cc}
 & y_1 \\
\boldsymbol{Y} = & \left|\begin{array}{c}
y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \\ y_6 \\ y_7 \\ y_8 \\ y_9 \\ y_{10} \\ y_{11} \\ y_{12} \\ y_{13} \\ y_{14} \\ y_{15} \\ y_{16} \\ y_{17} \\ y_{18} \\ y_{19} \\ y_{20}
\end{array}\right|
\end{array}. \tag{2.36}$$

Таблица 2.5

Величины радиусов сфер ядра плана, звёздных точек и количества опытов в центре плана в ротатабельных униформ-планах

Параметры плана	Количество факторов k								
	2	3	4	5	5	6	6	7	7
Ядро плана	2^2	2^3	2^4	2^5	2^{5-1}	2^6	2^{6-1}	2^7	2^{7-1}
$\rho_я$	1.41	1.73	2.00	2.24	2.27	2.45	2.45	2.64	2.64
α	1.41	1.68	2.00	2.38	2.00	2.83	2.38	3.36	2.83
n_0	5	6	7	10	6	15	9	21	14

Матрица ротатабельного плана второго порядка неортогональна, поскольку

$$\sum_{i=1}^{n} z^2_{i,j} \neq n; \qquad j=1,\ 2,\dots,\ k; \tag{2.37}$$

$$\sum_{i=1}^{n} z_{i,l} z_{i,m} \neq 0; \qquad l,\ m=1,\ 2,\dots,\ k; \quad l\neq m. \tag{2.38}$$

Вычисление коэффициентов уравнения регрессии

$$\hat{y} = b_0 + \sum_{j=1}^{k} b_j z_j + \sum_{\substack{l=1\\ m=2\\ l\neq m}}^{k} b_{l,m} z_l z_m + \sum_{j=1}^{k} b_{jj} z_j^2 \tag{2.39}$$

можно производить по формуле $\boldsymbol{B}=(\boldsymbol{Z}^{\mathrm{T}}\boldsymbol{Z})^{-1}(\boldsymbol{Z}^{\mathrm{T}}\boldsymbol{Y})$, предварительно дополнив матрицу $\boldsymbol{Z}$ соответствующими столбцами. Поскольку особенности матрицы $\boldsymbol{Z}^{\mathrm{T}}\boldsymbol{Z}$ для ротатабельного униформ-плана позволяют провести обращение этой матрицы в общем виде, можно получить формулы для расчёта коэффициентов уравнения регрессии и их дисперсий:

$$b_0 = \frac{A}{n} \cdot \left\{ 2\lambda_4^2 \cdot (k+2) \cdot \sum_{i=1}^{n} y_i - 2\lambda_4 C \cdot \sum_{j=1}^{k} \sum_{i=1}^{n} z_{ij}^2 y_i \right\}; \tag{2.40}$$

$$b_j = \frac{C}{n} \cdot \sum_{i=1}^{n} z_{ij} y_i; \quad j=1,\ 2,\dots,\ k; \tag{2.41}$$

$$b_{l,m} = \frac{C^2}{n\lambda_4} \cdot \sum_{i=1}^{n} z_{i,l} z_{i,m} y_i; \quad \begin{matrix} l=1,\ 2,\dots,\ k-1; \\ m=2,\ 3,\dots,\ k; \quad l\neq m; \end{matrix} \tag{2.42}$$

$$b_{jj} = \frac{A}{n}\left\{C^2\cdot\left[(k+2)\lambda_4 - k\right]\cdot\sum_{i=1}^{n} z_{ij}^2 y_i + C^2(1-\lambda_4)\cdot\sum_{j=1}^{k}\sum_{i=1}^{n} z_{ij}^2 y_i - \right.$$

$$\left. -2\lambda_4 C\cdot\sum_{i=1}^{n} y_i\right\};\quad j=1,\ 2,\ \ldots,\ k; \qquad (2.43)$$

$$s_{b_0}^2 = \frac{2A\lambda_4^2(k+2)}{n}\cdot s_{\text{оп}}^2; \qquad (2.44)$$

$$s_{b_j}^2 = \frac{C}{n}\cdot s_{\text{оп}}^2; \qquad (2.45)$$

$$s_{b_{l,m}}^2 = \frac{C^2}{\lambda_4 n}\cdot s_{\text{оп}}^2; \qquad (2.46)$$

$$s_{b_{jj}}^2 = \frac{A[(k+1)\lambda_4 - (k-1)]C^2}{n}\cdot s_{\text{оп}}^2, \qquad (2.47)$$

где

$$C = n\Big/\sum_{i=1}^{n} z_{ij}^2;\quad j\neq 0; \qquad (2.48)$$

$$\lambda_4 = \frac{kn}{(k+2)(n_я + n_\alpha)}; \qquad (2.49)$$

$$A = \frac{1}{2\lambda_4[(k+2)\lambda_4 - k]}. \qquad (2.50)$$

2.2.1. Проверка коэффициентов уравнения на значимость

Проверка каждого u-того коэффициента на *значимость* осуществляется путём сравнения опытного критерия Стьюдента с табличным:

$$t_{\text{оп}} = \frac{|b_u|}{s_{b_u}} \Leftrightarrow t_{\nu_{\text{оп}}}^{\alpha}; \qquad (2.51)$$

где табличное значение критерия Стьюдента берётся для числа степеней свободы дисперсии воспроизводимости $\nu_{\text{оп}}$ и уровня значимости α. Если опытное значение t-критерия попадает в критическую область (рис. 1.3), область отклонения нулевой гипотезы об отсутствии различия коэффициента b_u и его стандартного отклонения s_{bu}, коэффициент b_u значимо отличается от нуля. Риск от принятия такого решения не превышает уровня значимости α. Если опытное значение t-критерия меньше табличного, то с доверительной вероятностью P=1-α следует принять нулевую гипотезу об отсутствии различия коэффициента b_u и его стандартного отклонения s_{bu}, т.е. коэффициент b_u незначимо отличается от нуля. Если незначимым оказались один или более квадратичных эффектов, то после их исключения коэффициенты уравнения и их дисперсии необходимо пересчитать.

2.2.2. Проверка уравнения регрессии на адекватность

В научной и справочной литературе определение дисперсии воспроизводимости по результатам опытов в центре плана принято по умолчанию. И действительно, зачем ставить опыты в другой области факторного пространства, если в центре плана они обязательны по матрице плана. Но, поскольку отладка методики эксперимента требует постановки опытов на воспроизводимость перед собственно реализацией ротатабельного плана, вполне возможно, что к моменту проверки коэффициентов на значимость в распоряжении исследователя будет две выборки на воспроизводимость - в центре плана и в какой-то случайно выбранной вначале области факторного пространства. Если эта случайно выбранная вначале точка факторного пространства находится за пределами сферы с радиусом ρ_α, то обсуждать, по существу, нечего, а если внутри сферы, то у исследователя возможны альтернативы при проверке уравнения на адекватность. В отношении проверки коэффициентов на значимость сомнений быть не может: поскольку при вычислении коэффициентов используются результаты всех опытов ротатабельного плана, то при проверке коэффициентов на значимость должна быть использована дисперсия, вычисленная по результатам опытов в центре плана.

Иначе обстоит дело с проверкой уравнения на адекватность. В англо-американской литературе [10, 15] принято выделять сумму квадратов, определяющую неадекватность S_{LF} (Lack of Fit - *дословно*: недостаток соответствия), из остаточной суммы квадратов. Это значит, что если для проверки уравнения на адекватность используется дисперсия воспроизводимости опытов в центре плана, то сумма квадратов S_E (от Experiment - опыт), соответствующая этим опытам

$\{\Sigma(y_{0,i}-y_{0,cp})^2\}$, должна быть изъята из общей так называемой остаточной суммы квадратов S_R (от Remainder - остаток). Другими словами, сумму квадратов, вычисленную по всем опытам ротатабельного плана, следует разложить на две составляющих - сумму квадратов неадекватности (недостатка соответствия) и опытную сумму квадратов:

$$S_R = S_{LF} + S_E, \tag{2.52}$$

где

$$S_R = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \tag{2.53}$$

с числом степеней свободы $\nu_R = n-(k+2)(k+1)/2$, если все коэффициенты значимы, иначе $\nu_R = n-l$, где l - количество значимых коэффициентов;

$$S_{LF} = \sum_{i=1}^{n_я+n_\alpha} (y_i - \hat{y}_i)^2 \tag{2.54}$$

с числом степеней свободы $\nu_{LF} = n-(k+2)(k+1)/2-(n_0-1)$, если все коэффициенты значимы, иначе $\nu_{LF} = n-l-(n_0-1)$, где l - количество значимых коэффициентов, $n_я+n_\alpha$ - число периферийных точек плана, т.е. сумма опытов в ядре плана $n_я$ и в звёздных точках n_α

$$S_E = \sum_{i=1}^{n_0} (y_{0,i} - \bar{y}_0)^2 \tag{2.55}$$

с числом степеней свободы $\nu_E = n_0 - 1$.

Переходя к форме записи принятой в отечественной литературе, будем иметь:

остаточная дисперсия

$$s^2_{ост} = \frac{1}{n-l} \cdot \sum_{i=1}^{n} (y_i - \hat{y}_i)^2, \tag{2.56}$$

где $\nu_{ост} = n-l$; l - число связей, наложенных на выборку (общее число значимых коэффициентов в уравнении регрессии);

дисперсия адекватности

$$s^2_{ад} = \frac{1}{\nu_{ад}} \cdot \sum_{i=1}^{n_я+n_\alpha} (y_i - \hat{y}_i)^2, \tag{2.57}$$

где $\nu_{ад} = \nu_{ост} - \nu_{оп}$. Дисперсия воспроизводимости по формулам (1.16) и (1.17).

Опытный критерий Фишера:

$$F^{\text{оп}}_{\nu_{\text{ад}}\cdot\nu_{\text{оп}}} = \frac{S^2_{\text{ад}}}{S^2_{\text{оп}}}. \tag{2.58}$$

Опытное значение критерия Фишера $F^{\text{оп}}$ сравнивается с табличным значением F^{α} для чисел степеней свободы $\nu_{\text{ад}}$ и $\nu_{\text{оп}}$ и уровня значимости α. Если опытное значение критерия Фишера меньше табличного, то дисперсии $s^2_{\text{ад}}$ и $s^2_{\text{оп}}$ являются однородными, и, соответственно, уравнение регрессии адекватно. Если $F^{\text{оп}}>F^{\alpha}$, то дисперсии $s^2_{\text{ад}}$ и $s^2_{\text{оп}}$ считаются неоднородными, и полученное уравнение регрессии неадекватно.

Теперь вернёмся к вопросу об альтернативной проверке уравнения на адекватность. В принципе, можно за дисперсию адекватности принять остаточную дисперсию, вычисленную по всему массиву опытов ротатабельного плана, а за дисперсию воспроизводимости - дисперсию, вычисленную по опытам, поставленным в произвольной точке факторного пространства в пределах сферы ρ_{α}. Если в результате такой альтернативной проверки выводы об адекватности уравнения регрессии оказываются разными, то окончательное решение - на совести исследователя. Автор считает целесообразным выделять дисперсию неадекватности из остаточной дисперсии и при проверке уравнения на адекватность полученного по ортогональному плану, в том случае, если в качестве опытов на воспроизводимость используются опыты в центре плана.

2.2.3. Декодирование уравнения регрессии

Декодирование уравнения регрессии производится с помощью соотношения (2.26), используемого при кодировании переменных, аналогично ПФЭ. Следует также напомнить, что бывают случаи незначимости коэффициента b_j и значимости, например, b_{jj}. Согласно требованиям регрессионного анализа незначимый коэффициент b_j необходимо исключить, а значимый коэффициент b_{jj} оставить, но при декодировании уравнения коэффициент при j-том факторе появится и j-тый фактор как бы возвращается в уравнении регрессии. Вопрос с действительной значимостью j-того фактора остаётся открытым.

2.2.4. Пример ротатабельного планирования

При стендовом испытании бурового инструмента новой конструкции было произведено ротатабельное центральное композиционное планирование. При этом исследователей интересовало влияние скорости вращения инструмента и нагрузки на скорость проходки.

Обозначения и пределы изменения переменных:

x_1 – n, скорость вращения инструмента, 125÷495 об/мин;

x_2 – P, нагрузка на инструмент, 7,1÷10,3 кН;

y – $w_м$, механическая скорость бурения, м/ч.

Требуется получить уравнение регрессии вида

$$w_м = b_0 + b_1 n + b_2 P + b_{1.2} nP + b_{11} n^2 + b_{22} P^2.$$

Кодирование переменных

Координаты центра плана и интервал варьирования переменных вычисляются по формулам

$$x_j^0 = \frac{x_j^{max} + x_j^{min}}{2}; \qquad \Delta x_j = \frac{x_j^{max} - x_j^{min}}{2}.$$

Количество экспериментов в центре плана n_0 и величина звёздного плеча α, соответствующие числу факторов $k=2$, находим по табл. 2.5. См. также (1.16) и (1.17). Результаты кодирования представлены в табл. 2.6.

Матрица двухфакторного плана, результаты экспериментов и результаты расчётов по искомому уравнению (см. ниже) представлены в табл. 2.7.

Вычисление коэффициентов b_0, b_j, $b_{1.m}$ и b_{jj} производится по формулам (2.40), (2.41), (2.42), (2.43), результаты представлены в табл. 2.8.

Дисперсия воспроизводимости определяется по формулам (1.16) и (1.17):

$$\bar{y}_0 = \frac{1}{n_0} \cdot \sum_{i=1}^{n_0} y_{0,i} = \frac{5,32}{5} = 1,064;$$

$$s_{оп}^2 = \frac{1}{(n_0 - 1)} \sum_{i=1}^{n_0} (y_{0,i} - \bar{y}_0)^2 = \frac{0,00132}{4} = 0,00033.$$

Таблица 2.6.

Результаты кодирования переменных

Компоненты ротатабельного центрального композиционного плана	Значения факторов			
	Скорость вращения		Нагрузка на инструмент	
	Физические	Код. z_1	Физические	Код. z_2
Центр плана	310	0	28,7	0
Интервал варьирования	185	1	1,6	1
Звёздное плечо	261,6	1,414	2,26	1,414
Нижний уровень	125	−1	7,1	−1
Верхний уровень	495	+1	10,3	+1

Таблица 2.7.

Матрица плана, результаты экспериментов и расчётов по искомому уравнению регрессии

z_0	z_1	z_2	$y_{1,\text{эксп}}$, м/ч	$y_{1,\text{расч}}$, м/ч
+1	+1	+1	1,58	1,611
+1	−1	+1	0,71	0,733
+1	+1	−1	1,16	1,177
+1	−1	−1	0,068	0,0773
+1	+1.414	0	1,62	1,594
+1	−1.414	0	0,21	0,1954
+1	0	+1.414	1,32	1,290
+1	0	−1.414	0,53	0,5195
+1	0	0	1,05	1,064
+1	0	0	1,08	1,064
+1	0	0	1,04	1,064
+1	0	0	1,08	1,064
+1	0	0	1,07	1,064

Таблица 2.8.

Коэффициенты уравнения с кодированными переменными, значения экспериментального критерия Стьюдента и коэффициенты уравнения с физическими переменными

Коэф. b_j	Коэффициенты кодированного уравнения	Эксп. критерий Стьюдента	Коэффициенты декодированного уравнения
b_0	1,064	131	-4,3434
b_1	0,4945	77	0,005837
b_2	0,2724	42,4	0,7696
$b_{1,2}$	-0,0555	6,11	-0,0001875
$b_{1,1}$	-0,08463	12,3	$-0,2473\cdot10^{-5}$
$b_{2,2}$	-0,07963	11,6	-0,0311

Дисперсии коэффициентов уравнения регрессии определяются по формулам (2.44), (2.45), (2.46) и (2.47):

$$s^2_{b_0} = 0,66\cdot10^{-4}; \qquad s^2_{b_j} = 0,4125\cdot10^{-4};$$

$$s^2_{b_{l,m}} = 0,825\cdot10^{-4}; \qquad s^2_{b_{j,j}} = 0,4744\cdot10^{-4}.$$

Соответствующие квадратичные отклонения коэффициентов уравнения регрессии представлены ниже:

$$s_{b_0}=\pm0,008124; \quad s_{b_j}=\pm0,006423; \quad s_{b_{l,m}}=\pm0,009083; \quad s_{b_{j,j}}=\pm0,006887.$$

Вычисленные по формуле (2.19) опытные критерии Стьюдента для каждого коэффициента уравнения регрессии представлены в табл. 2.8. Для уровня значимости α=0,05 и числа степеней свободы дисперсии воспроизводимости ν=4 табличное значение двустороннего критерия Стьюдента t=2,78. Результаты сравнения опытных критериев Стьюдента с табличным значением показывают, что все опытные значения t-критериев попадают в критическую область (рис. 1.3), область отклонения нулевой гипотезы об отсутствии различия коэффициента b_j и его стандартного отклонения s_{bj}. Таким образом, с доверительной вероятностью P=0,95 все коэффициенты значимы. Уравнение регрессии с кодированными переменными имеет вид

$$y=b_0+b_1 z_1+b_2 z_2+b_{1,2} z_1 z_2+b_{11} z^2{}_1+b_{22} z^2{}_2.$$

Проверку уравнения регрессии на адекватность осуществим по методике, принятой в англо-американской литературе. Дисперсию адекватности вычислим по формуле (2.57):

$$s^2_{\text{ад}}=\frac{1}{\nu_{\text{ад}}}\cdot\sum_{i=1}^{n_я+n_\alpha}(y_1-\hat{y}_1)^2=\frac{0,00379}{3}=0,001263,$$

где $\nu_{\text{ад}}=n-l-(n_0-1)=13-6-(5-1)=3$. Расчетное значение функции отклика можно получить по уравнениям (1.4), (2.16) или (2.39).

Опытный критерий Фишера вычисляем по формуле (2.24):

$$F^{\text{оп}}_{\nu_{\text{ад}},\nu_{\text{оп}}}=\frac{s^2_{\text{ад}}}{s^2_{\text{оп}}}=\frac{0,001263}{0,00033}=3,83,$$

где $\nu_{\text{ад}}$=3, $\nu_{\text{оп}}$=4.

Опытный критерий Фишера сравниваем с табличным (табл. П.2):

$$F^{0,05}_{3,4}=6,6>F^{\text{оп}}_{3,4}=3,83.$$

Опытный критерий Фишера не попадает в критическую область (рис. 1.2, он меньше табличного значения, взятого для уровня значимости α=0,05. Это значит, что с доверительной вероятностью P=0,95 нулевая гипотеза об однородности дисперсий $s^2{}_{\text{ад}}$ и $s^2{}_{\text{оп}}$ не противоречит результатам проверки, т.е. уравнение регрессии адекватно.

Декодирование уравнения регрессии

Декодирование уравнения регрессии производится с помощью соотношения (2.26), используемого при кодировании переменных:

$$y=b_0+b_1\left(\frac{x_1-x_1^0}{\Delta x_1}\right)+b_2\left(\frac{x_2-x_2^0}{\Delta x_2}\right)+b_{1,2}\left(\frac{x_1-x_1^0}{\Delta x_1}\right)\left(\frac{x_2-x_2^0}{\Delta x_2}\right)+b_{11}\left(\frac{x_1-x_1^0}{\Delta x_1}\right)^2+b_{22}\left(\frac{x_2-x_2^0}{\Delta x_2}\right)^2.$$

Раскрывая скобки, приводя подобные члены и подставляя численные значения коэффициентов уравнения, координат центра плана и интервалов варьирования факторов, получим

$$y=-4,34+0,00584x_1+0,77x_2-0,000188x_1x_2-0,247\cdot10^{-5}x_1^2-0,0311x_2^2,$$

переходя к обозначениям физических переменных

$$w_м=-4,34+0,00584n+0,77P-0,000188nP-0,247\cdot10^{-5}n^2-0,0311P^2.$$

Полученное уравнение описывает зависимость механической скорости бурения от скорости вращения инструмента и нагрузки.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Математический энциклопедический словарь. М.: Сов. энциклопедия, 1988.- 847 с.
2. *Ахназарова С.Л., Кафаров В.В.* Методы оптимизации эксперимента в химической технологии: Учеб. пособ. для хим.-технол. спец. вузов. 2-е изд., перераб. и доп.- М.: Высш. шк., 1985.- 327 с.
3. *Большев Л.Н., Смирнов Н.В.* Таблицы математической статистики.- 3 изд. М., 1983.
4. *Ганджумян Р.А.* Математическая статистика в разведочном бурении: Справ. пособ.- М.: Недра, 1990.- 218 с.
5. *Кендалл М.Дж., Стьюарт А.* Статистические выводы и связи: Пер. с англ.- М.: Наука, - 1973, - 900 с.
6. *Кендалл М.Дж., Стьюарт А.* Теория распределений: Пер с англ.- М.: Наука, 1966.- 588 с.
7. *Крамер Г.* Математические методы статистики: Пер. с англ. Изд. 2-е.- М.: Мир, 1975. - 648 с.
8. *Линнин Ю.В.* Метод наименьших квадратов и основы математико-статистической теории обработки наблюдений. Изд. 2-е.- М., 1962.
9. *Львовский Е.Н.* Статистические методы построения эмпирических формул: Учеб. пособ.- М.: Высш. шк., 1982.- 224 с.
10. *Налимов В.В., Чернова Н.А.* Статистические методы планирования экстремальных экспериментов. М.: Наука, 1965. 340 с.
11. *Саутин С.Н.* Планирование эксперимента в химии и химической технологии. Л.: Химия, 1975. 48 с.
12. *Хьютсон А.* Дисперсионный анализ.- М.: Статистика, 1971.
13. *Шеффе Д.* Дисперсионный анализ.- М.: Физматгиз, 1963.
14. *Finney D.J.* The fraktional replication of factorial experiments// Annals of Eugenics. 1945. 12. №4. 291.
15. *Box G.E.P., Hunter J.S.* Multifactor experimental disigns for exploring response surfaces// Annals of mathematical statistics. 1957. 28. №1. 195.
16. *Дэвис Дж.С.* Статистический анализ данных в геологии: В 2 кн./Пер. с англ. В.А.Голубевой: Под ред. Д.А.Родионова.- М.: Недра, 1990.- 427 с.
17. *Цивинский Д.Н.* Применение метода полного факторного эксперимента в нефтегазовом деле: Учеб. пособ. / Самар. гос. техн. ун-т., Самара, 2001, 87 с.
18. *Цивинский Д.Н.* Применение метода пассивного эксперимента в нефтегазовом деле: Учеб. пособ. / Самар. гос. техн. ун-т. Самара, 2001, 84 с.

Таблица П.1.

Критические значения одностороннего критерия Стьюдента при ν степенях свободы и заданном уровне значимости α

Число степеней свободы ν	Уровень значимости α					
	0,1	0,05	0,025	0,01	0,005	0,001
1	3,078	6,314	12,706	31,821	63,657	318,310
2	1,886	2,920	4,303	6,965	9,925	22,327
3	1,638	2,353	3,182	4,541	5,841	10,215
4	1,533	2,132	2,776	3,747	4,604	7,173
5	1,476	2,015	2,571	3,365	4,032	5,893
6	1,440	1,943	2,447	3,143	3,707	5,208
7	1,415	1,895	2,365	2,998	3,499	4,785
8	1,397	1,860	2,306	2,896	3,355	4,501
9	1,383	1,833	2,262	2,821	3,250	4,297
10	1,372	1,812	2,228	2,764	3,169	4,144
11	1,363	1,796	2,201	2,718	3,105	4.025
12	1,356	1,782	2,179	2,681	3,055	3,930
13	1,350	1,771	2,160	2,650	3,012	3,852
14	1,345	1,761	2,145	2,624	2,977	3,787
15	1,341	1,753	2,131	2,602	2,947	3,733
16	1,377	1,746	2,120	2,583	2,921	3,686
17	1,333	1,740	2,110	2,567	2,898	3,646
18	1,330	1,734	2,101	2,552	2,878	3,610
19	1,328	1,729	2,093	2,539	2,861	3,579
20	1,325	1,725	2,086	2,528	2,845	3,552
21	1,323	1,721	2,080	2,518	2,831	3,527
22	1,321	1,717	2,074	2,508	2,819	3,505
23	1,319	1,714	2,069	2,500	2,807	3,485
24	1,318	1,711	2,064	2,492	2,797	3,467
25	1,316	1,708	2,060	2,485	2,787	3,450
26	1,315	1,706	2,056	2,479	2,779	3,435
27	1,314	1,703	2,052	2,473	2,771	3,421
28	1,313	1,701	2,048	2,467	2,763	3,408
29	1,311	1,699	2,045	2,462	2,756	3,396
30	1,310	1,697	2,042	2,457	2,750	3,385
40	1,303	1,684	2,021	2,423	2,704	3,307
60	1,296	1,671	2,000	2,390	2,660	2,232
120	1,289	1,658	1,980	2,358	2,617	3,160
∞	1,282	1,645	1,960	2,326	2,576	3,090

Таблица П.2

Критические значения распределения Фишера F^{α} для уровня значимости α=0,05

ν_2	Число степеней свободы числителя ν_1																		
	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
1	161,4	199,5	215,7	224,6	230,2	234,0	236,8	238,9	240,5	241,9	243,9	245,9	248,0	249,1	250,1	251,1	252,2	253,3	254,3
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	19,45	19,46	19,47	19,48	19,49	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	8,64	8,62	8,59	8,57	8,55	8,53
4	7,71	6,94	6,59	6,39	6,26	6,16	6,09	6,04	6.00	5,96	5,91	5,86	5,80	5,77	5,75	5,72	5,69	5,66	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	4,53	4,50	4,46	4,43	4,40	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	3,84	3,81	3,77	3,74	3,70	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	3,41	3,38	3,34	3,30	3,27	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	3,12	3.08	3,04	3,01	2,97	2,93
9	5,12	4,26	3,86	3,63	3,48	3,37	3,29	3,23	3,18	3,14	3,07	3,01	2,94	2,90	2,86	2,83	2,79	2,75	2,71
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	2,74	2,70	2,66	2,62	2,58	2,54
11	4,84	3,98	3,59	3,36	3,20	3,09	3,01	2,95	2,90	2,85	2,79	2,72	2,65	2,61	2,57	2,53	2,49	2,45	2,40
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	2,51	2,47	2,43	2,38	2,34	2,30
13	4,67	3,81	3,41	3,18	3,03	2,92	2,83	2,77	2,71	2,67	2,60	2,53	2.46	2,42	2,38	2,34	2,30	2,25	2,21
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	2,35	2,31	2,27	2,22	2,18	2,13
15	4,54	3,68	3,29	3,06	2,90	2,79	2,71	2,64	2,59	2,54	2,48	2,40	2,33	2,29	2,25	2,20	2,16	2,11	2,07
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	2,24	2,19	2,15	2,11	2,06	2,01
17	4,45	3,59	3,20	2,96	2,81	2,70	2,61	2,55	2,49	2,45	2,38	2,31	2,23	2,19	2.15	2,10	2,06	2,01	1,96
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	2,15	2,11	2,06	2,02	1,97	1,92
19	4,38	3,52	3,13	2,90	2,74	2,63	2,54	2,48	2,42	2,38	2,31	2,23	2,16	2,11	2,07	2,03	1,98	1,93	1,88
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	2,08	2,04	1,99	1,95	1,90	1,84
21	4,32	3,47	3,07	2,84	2,68	2,57	2,49	2,42	2,37	2,32	2,25	2,18	2,10	2,05	2,01	1,96	1,92	1,87	1,81
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	2,03	1,98	1,94	1,89	1,84	1,78
23	4,28	3,42	3,03	2,80	2,64	2,53	2,44	2,37	2,32	2,27	2,20	2,13	2,05	2,01	1,96	1,91	1,86	1,81	1,76
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	1,98	1,94	1,89	1,84	1,79	1,73
25	4,24	3,39	2,99	2,76	2,60	2,49	2,40	2,34	2,28	2,24	2,16	2,09	2,01	1,96	1,92	1,87	1,82	1,77	1,71
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	1,95	1,90	1,85	1,80	1,75	1,69
27	4,21	3,35	2,96	2,73	2,57	2,46	2,37	2,31	2,25	2,20	2,13	2,06	1,97	1,93	1,88	1,84	1,79	1,73	1,67
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	1,91	1,87	1,82	1,77	1,71	1,65
29	4,18	3,33	2,93	2,70	2,55	2,43	2,35	2,28	2,22	2,18	2,10	2,03	1,94	1,90	1,85	1,81	1,75	1,70	1,64
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	1,89	1,84	1,79	1,74	1,68	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	1,79	1,74	1,69	1,64	1,58	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	1,70	1,65	1,59	1,53	1,47	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	1,61	1,55	1,50	1,43	1,35	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	1,52	1,46	1,39	1,32	1,22	1,00

Таблица П.3

Критические значения распределения Фишера F^{α} для уровня значимости α=0,025

ν_2	Число степеней свободы числителя, ν_1																		
	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
1	4052	499,5	5403	5625	5764	5859	5928	5981	6022	6056	6106	6157	6209	6235	6261	6287	6313	6339	6366
2	98,50	99,00	99,17	99,25	99,30	99,33	99,36	99,37	99,39	99,40	99,42	99,43	99,45	99,46	99,47	99,47	99,48	99,49	99,50
3	34,12	30,82	29,46	28 71	28,24	27,91	27,67	27,49	27,35	27,23	27,05	26,87	26,69	26,60	26,50	26,41	26,32	26,22	26,13
4	21,20	18,00	16,69	15,98	15,52	15,21	14,98	14,80	14,66	14,55	14,37	14,20	14,02	13,93	13,84	13,75	13,65	13,56	13,46
5	16,26	13,27	12,06	11,39	10,97	10,67	10,46	10,29	10,16	10,05	9,89	9,72	9,55	9,47	9,38	9,29	9,20	9,11	9,02
6	13,75	10,92	9,78	9,15	8,75	8,47	8,26	8,10	7,98	7,87	7,72	7,56	7,40	7,31	7,23	7,14	7,06	6,97	6,88
7	12,25	9,55	8,45	7,85	7,46	7,19	6,99	6,84	6,72	6,62	6,47	6,31	6,16	6,07	5,99	5,91	5,82	5,74	5,65
8	11,26	8,65	7,59	7,01	6,63	6,37	6,18	6,03	5,91	5,81	5,67	5,52	5,36	5,28	5,20	5,12	5,03	4,95	4,86
9	10,56	8,02	6,99	6,42	6,06	5,80	5,61	5,47	5,35	5,26	5,11	4,96	4,81	4,73	4,65	4,57	4,48	4,40	4,31
10	10,04	7,56	6,55	5,99	5,64	5,39	5,20	5,06	4,94	4,85	4,71	4,56	4,41	4,33	4,25	4,17	4,08	4,00	3,91
11	9,65	7,21	6,22	5,67	5,32	5,07	4,89	4,74	4,63	4,54	4,40	4,25	4,10	4,02	3,94	3,86	3,78	3,69	3,60
12	9,33	6,93	5,95	5,41	5,06	4,82	4,64	4,50	4,39	4,30	4,16	4,01	3,86	3,78	3,70	3,62	3,54	3,45	3,36
13	9,07	6,70	5,74	5,21	4,86	4,62	4,44	4,30	4,19	4,10	3,96	3,82	3,66	3,59	3,51	3,43	3,34	3,25	3,17
14	8,86	6,51	5,56	5,04	4,69	4,46	4,28	4,14	4,03	3,94	3,80	3,66	3,51	3,43	3,35	3,27	3,18	3,09	3,00
15	8,68	6,36	5,42	4,89	4,56	4,32	4,14	4,00	3,89	3,80	3,67	3,52	3,37	3,29	3,21	3,13	3,05	2,96	2,87
16	8,53	6,23	5,29	4,77	4,44	4,20	4,03	3,89	3,78	3,69	3,55	3,41	3,26	3,18	3,10	3,02	2,93	2,84	2,75
17	8,40	6,11	5,18	4,67	4,34	4,10	3,93	3,79	3,68	3,59	3,46	3,31	3,16	3,08	3,00	2,92	2,83	2,75	2,65
18	8,29	6,01	5,09	4,58	4,25	4,01	3,84	3,71	3,60	3,51	3,37	3,23	3,08	3,00	2,92	2,84	2,75	2,66	2,57
19	8,18	5,93	5,01	4,50	4,17	3,94	3,77	3,63	3,52	3,43	3,30	3,15	3,00	2,92	2,84	2,76	2,67	2,58	2,49
20	8,10	5,85	4,94	4,43	4,10	3,87	3,70	3,56	3,46	3,37	3,23	3,09	2,94	2,86	2,78	2,69	2,61	2,52	2,42
21	8,02	5,78	4,87	4,37	4,04	3,81	3,64	3,51	3,40	3,31	3,17	3,03	2,88	2,80	2,72	2,64	2,55	2,46	2,36
22	7,95	5,72	4,82	4,31	3,99	3,76	3,59	3,45	3,35	3,26	3,12	2,98	2,83	2,75	2,67	2,58	2,50	2,40	2,31
23	7,88	5,66	4,76	4,26	3,94	3,71	3,54	3,41	3,30	3,21	3,07	2,93	2,78	2,70	2,62	2,54	2,45	2,35	2,26
24	7,82	5,61	4,72	4,22	3,90	3,67	3,50	3,36	3,26	3,17	3,03	2,89	2,74	2,66	2,58	2,49	2,40	2,31	2,21
25	7,77	5,57	4,68	4,18	3,85	3,63	3,46	3,32	3,22	3,13	2,99	2,85	2,70	2,62	2,54	2,45	2,36	2,27	2,17
26	7,72	5,53	4,64	4,14	3,82	3,59	3,42	3,29	3,18	3,09	2,96	2,81	2,66	2,58	2,50	2,42	2,33	2,23	2,13
27	7,68	5,49	4,60	4,11	3,78	3,56	3,39	3,26	3,15	3,06	2,93	2,78	2,63	2,55	2,47	2,38	2,29	2,20	2,10
28	7,64	5,45	4,57	4,07	3,75	3,53	3,36	3,23	3,12	3,03	2,90	2,75	2,60	2,52	2,44	2,35	2,26	2,17	2,06
29	7,60	5.42	4,54	4,04	3,73	3,50	3,33	3,20	3,09	3,00	2,87	2,73	2,57	2,49	2,41	2,33	2,23	2,14	2,03
30	7,56	5,39	4,51	4,02	3,70	3,47	3,30	3,17	3,07	2,98	2,84	2,70	2,55	2,47	2,39	2,30	2,21	2,11	2,01
40	7,31	5,18	4,31	3,83	3,51	3,29	3,12	2,99	2,89	2,80	2,66	2,52	2,37	2,29	2,20	2,11	2,02	1,92	1,80
60	7,08	4,98	4,13	3,65	3,34	3,12	2,95	2,82	2,72	2,63	2,50	2,35	2,20	2,12	2,03	1,94	1,84	1,73	1,60
120	6,85	4,79	3,95	3,48	3,17	2,96	2,79	2,66	2,56	2,47	2,34	2,19	2,03	1,95	1,86	1,76	1,66	1,53	1,38
∞	6,63	4,61	3,78	3,32	3,02	2,80	2,64	2,51	2,41	2,32	2,18	2,04	1,88	1,79	1,70	1,59	1,47	1,32	1,00

Таблица П.4

Критические значения распределения Фишера F^{α} для уровня значимости α=0,001

ν_2	Число степеней свободы числителя, ν_1																		
	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
1	4053*	5000*	5404*	5625*	5764*	5859*	5929*	5981*	6023*	6056*	6107*	6158*	6209*	6235*	6261*	6287*	6313*	6340*	6366*
2	998,5	999,0	999,2	999,2	999,3	999,3	999,4	999,4	999,4	999,4	999,4	999,4	999,4	999,5	999,5	999,5	999,5	999,5	999,5
3	167,0	149,	141,1	137,1	134,6	132,8	131,6	130,6	129,9	129,2	128,3	127,4	126,4	125,9	125,4	125,0	124,5	124,0	123,5
4	74,14	61,25	56,18	53,44	51,71	50,53	49,66	49,00	48,47	48,05	47,41	46,76	46,10	45,77	45,43	45,09	44,75	44,40	44,05
5	47,18	37,12	33,20	31,09	29,75	28,84	28,16	27,64	27,24	26,92	26,42	25,91	25,39	25,14	24,87	24,60	24,33	24,06	23,79
6	35,51	27,00	23,70	21,92	20,81	20,03	19,46	19,03	18,69	18,41	17,99	17,56	17,12	16,89	16,67	16,44	16,21	15,99	15,75
7	29,25	21,69	18,77	17,19	16,21	15,52	15,02	14,63	14,33	14,08	13,71	13,32	12,93	12,73	12,53	12,33	12,12	11,91	11,70
8	25,42	18,49	15,83	14,39	13,49	12,86	12,40	12,04	11,77	11,54	11,19	10,84	10,48	10,30	10,11	9,92	9,73	9,53	9,33
9	22,86	16,39	13,90	12,56	11,71	11,13	10,70	10,37	10,11	9,89	9,57	9,24	8,90	8,72	8,55	8,37	8,19	8,00	7,81
10	21,04	14,91	12,55	11,28	10,48	9,92	9,52	9,20	8,96	8,75	8,45	8,13	7,80	7,64	7,47	7,30	7,12	6,94	6,76
11	19,69	13,81	11,56	10,35	9,58	9,05	8,66	8,35	8,12	7,92	7,63	7,32	7,01	6,85	6,68	6,52	6,35	6,17	6,00
12	18,64	12,97	10,80	9,63	8,89	8,38	8,00	7,71	7,48	7,29	7,00	6,71	6,40	6,25	6,09	5,93	5,76	5,59	5,42
13	17,81	12,31	10,21	9,07	8,35	7,86	7,49	7,21	6,98	6,80	6,52	6,23	5,93	5,78	5,63	5,47	5,30	5,14	4,97
14	17,14	11,78	9,73	8,62	7,92	7,43	7,08	6,80	6,58	6,40	6,13	5,85	5,56	5,41	5,25	5,10	4,94	4,77	4,60
15	16,59	11,34	9,34	8,25	7,57	7,09	6,74	6,47	6,26	6,08	5,81	5,54	5,25	5,10	4,95	4,80	4,64	4,47	4,31
16	16,12	10,97	9,00	7,94	7,27	6,81	6,46	6,19	5,98	5,81	5,55	5,27	4,99	4,85	4,70	4,54	4,39	4,23	4,06
17	15,72	10,66	8,73	7,68	7,02	6,56	6,22	5,96	5,75	5,58	5,32	5,05	4,78	4,63	4,48	4,33	4,18	4,02	3,85
18	15,38	10,39	8,49	7,46	6,81	6,35	6,02	5,76	5,56	5,39	5,13	4,87	4,59	4,45	4,30	4,15	4,00	3,84	3,67
19	15,08	10,16	8,28	7,26	6,62	6,18	5,85	5,59	5,39	5,22	4,97	4,70	4,43	4,29	4,14	3,99	3,84	3,68	3,51
20	14,82	9,95	8,10	7,10	6,46	6,02	5,69	5,44	5,24	5,08	4,82	4,56	4,29	4,15	4,00	3,86	3,70	3,54	3,38
21	14,59	9,77	7,94	6,95	6,32	5,88	5,56	5,31	5,11	4,95	4,70	4,44	4,17	4,03	3,88	3,74	3,58	3,42	3,26
22	14,38	9,61	7,80	6,81	6,19	5,76	5,44	5,19	4,99	4,83	4,58	4,33	4,06	3,92	3,78	3,63	3,48	3,32	3,15
23	14,19	9,47	7,67	6,69	6,08	5,65	5,33	5,09	4,89	4,73	4,48	4,23	3,96	3,82	3,68	3,53	3,38	3,22	3,05
24	14,03	9,34	7,55	6,59	5,98	5,55	5,23	4,99	4,80	4,64	4,39	4,14	3,87	3,74	3,59	3,45	3,29	3,14	2,97
25	13,88	9,22	7,45	6,49	5,88	5,46	5,15	4,91	4,71	4,56	4,31	4,06	3,79	3,66	3,52	3,37	3,22	3,06	2,89
26	13,74	9,12	7,36	6,41	5,80	5,38	5,07	4,83	4,64	4,48	4,24	3,99	3,72	3,59	3,44	3,30	3,15	2,99	2,82
27	13,61	9,02	7,27	6,33	5,73	5,31	5,00	4,76	4,57	4,41	4,17	3,92	3,66	3,52	3,38	3,23	3,08	2,92	2,75
28	13,50	8,93	7,19	6,25	5,66	5,24	4,93	4,69	4,50	4,35	4,11	3,86	3,60	3,46	3,32	3,18	3,02	2,86	2,69
29	13,39	8,85	7,12	6,19	5,59	5,18	4,87	4,64	4,45	4,29	4,05	3,80	3,54	3,41	3,27	3,12	2,97	2,81	2,64
30	13,29	8,77	7,05	6,12	5,53	5,12	4,82	4,58	4,39	4,24	4,00	3,75	3,49	3,36	3,22	3,07	2,92	2,76	2,59
40	12,61	8,25	6,60	5,70	5,13	4,73	4,44	4,21	4,02	3,87	3,64	3,40	3,15	3,01	2,87	2,73	2,57	2,41	2,23
60	11,97	7,76	6,17	5,31	4,76	4,37	4,09	3,87	3,69	3,54	3,31	3,08	2,83	2,69	2,55	2,41	2,25	2,08	1,89
120	11,38	7,32	5,79	4,95	4,42	4,04	3,77	3,55	3,38	3,24	3,02	2,78	2,53	2,40	2,26	2,11	1,95	1,76	1,54
∞	10,83	6,91	5,42	4,62	4,10	3,74	3,47	3,27	3,10	2,96	2,74	2,51	2,27	2,13	1,99	1,84	1,66	1,45	1,00

* Эти значения надо умножить на 100

ОГЛАВЛЕНИЕ

Цивинский Дмитрий Николаевич

Применение метода центрального композиционного планирования в нефтегазовом деле

Редактор *С.И.Костерина*
Технический редактор *Г.Н.Шанькова*

Л.Р. №020595 от 09.07.97.
Подп. в печать 01.06.01.
Формат 60*84 1/16. Бум. типогр. N2.
Печать офсетная,
Усл.кр.-отт. 5,35. Уч.-изд.л. 5,07.
Тираж 170 экз. С-361.

Самарский государственный технический университет
443100, г. Самара, ул. Молодогвардейская, д. 244, Главный корпус.