



МИНИСТЕРСТВО
ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ

Государственное образовательное учреждение высшего
профессионального образования

САМАРСКИЙ ГОСУДАРСТВЕННЫЙ
ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ

Д.Н. Цивинский

ПРИМЕНЕНИЕ МЕТОДА ПОЛНОГО ФАКТОРНОГО ЭКСПЕРИМЕНТА 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»

УДК 519.22:681.3(076.5)

Применение метода полного факторного эксперимента в нефтегазовом деле: Учеб. пособ. / Д.Н. Цивинский; Самар. гос. техн. ун-т. Самара, 2002. 87 с.

Рассматривается применение метода полного факторного эксперимента к решению технологических задач строительства скважин: задачи поиска оптимальных условий проводки скважины, получения математической модели процесса без исследования механизма или описания вида "свойство=f(состава)" и др. Описаны теоретические основы кодирования факторов, построения матрицы плана, вычисления коэффициентов уравнения и проведения регрессионного анализа. В достаточном объёме рассмотрен вопрос проверки статистических гипотез. Для облегчения понимания подробно рассмотрены более 30 специальных терминов, причём определение каждого термина представлено отдельной статьёй. Пример расчёта иллюстрирует все этапы получения уравнения множественной регрессии и возможные действия в случае неопределённой значимости коэффициентов уравнения.

Предназначено для самостоятельной работы студентов очной и заочной форм обучения, могут быть полезны аспирантам и инженерам при практической обработке промышленных данных.

ISBN 5-7964-0235-8.

Табл. 6. Ил. 7. Библиогр. 20 назв.

Печатается по решению редакционно-издательского совета Самарского государственного технического университета

Рецензенты: В.А.Акулов,
В.К.Давыдов

© Д.Н.Цивинский, 2002.
© Самарский государственный
технический университет, 2002.

ISBN 5-7964-0235-8

Условные обозначения

Условное обозначение	Физическая величина
B	Формальный параметр или вектор параметров
b_j	Оценка j -того коэффициента уравнения регрессии (оценка генерального параметра β_j)
D	Детерминант матрицы
DX	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)
$S^2_{ад}$	Дисперсия адекватности
$S^2_{он}$	Дисперсия воспроизводимости

Условное обозначение	Физическая величина
s^2_x, s^2_y	Выборочные дисперсии (несмещённые оценки генеральных параметров, σ^2_x, σ^2_y)
t	Критерий Стьюдента
X, Y	Случайные величины
X	Матрица независимых переменных x_{1j}
\bar{x}, \bar{y}	Выборочные средние значения переменных (оценки математических ожиданий M_x, M_y)
x	Независимая переменная величина (фактор)
Y	Матрица-столбец функции отклика y_1
y	Функция отклика
\hat{y}	Значение функции отклика, рассчитанное по уравнению регрессии
Z	Матрица плана экспериментов
z	Кодированное значение переменной величины
α	Уровень значимости
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
κ	Неизвестный генеральный параметр
M_x, M_y	Математические ожидания случайных величин X, Y
μ_2	Центральный момент второго порядка
ν	Число степеней свободы
σ^2_x, σ^2_y	Генеральные дисперсии случайных величин X, Y
σ_x, σ_y	Генеральные квадратичные отклонения
v	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, 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 (структурная вязкость, напряжение сдвига, водоотдача и др.) с составом. Другой пример: при оптимизации технологического процесса очистки нефти найти такое сочетание температуры, давления, концентраций реагентов, структуры потоков и других значений, чтобы получить нефть требуемого качества.

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 практичес-

ких 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 объем 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.

1. ПОСТАНОВКА ЗАДАЧИ

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

$$y=f(x_1, x_2, \dots, x_k), \quad (1.1)$$

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, называется поверхностью отклика (рис. 1.1).

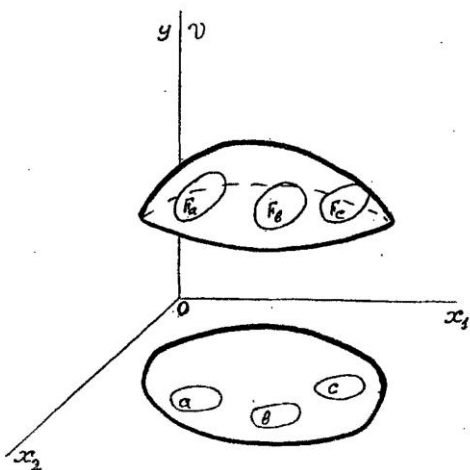


Рис. 1.1. Гипотетическая зависимость $U=f(x_1, x_2)$ в системе координат (x_1, x_2, U) . Возможные экспериментальные значения функции отклика $y=f(x_1, x_2)$ на рисунке не показаны.

Термин "функция отклика" принят для того, чтобы отличать функции, имеющие аналитическое выражение, 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 та-

ком случае аналитическое выражение функции отклика неизвестно, и её представляют полиномом:

$$v = \beta_0 + \sum_{j=1}^k \beta_j x_j + \sum_{\substack{l=1 \\ m=2 \\ l \neq m}}^k \beta_{l,m} x_l x_m + \sum_{j=1}^k \beta_{jj} x_j^2 \dots \quad (1.2)$$

где v - характеристика процесса, подлежащая оптимизации или регулированию; $\beta_0, \beta_j, \beta_{l,m}, \beta_{jj}, \dots$ - коэффициенты уравнения регрессии. Разложение неизвестной функции в степенной ряд эквивалентно представлению её рядом Тейлора:

$$\begin{aligned} \beta_1 &= \frac{\partial f}{\partial x_1}; \quad \beta_2 = \frac{\partial f}{\partial x_2}; \quad \dots; \quad \beta_k = \frac{\partial f}{\partial x_k}; \\ \beta_{1,2} &= \frac{\partial^2 f}{\partial x_1 \partial x_2}; \quad \beta_{2,3} = \frac{\partial^2 f}{\partial x_2 \partial x_3}; \quad \dots; \\ \beta_{1,1} &= \frac{1}{2} \cdot \frac{\partial^2 f}{\partial x^2}; \quad \beta_{2,2} = \frac{1}{2} \cdot \frac{\partial^2 f}{\partial x^2}; \quad \dots \end{aligned} \quad (1.3)$$

Очевидно, что для определения коэффициентов уравнения регрессии необходимо поставить некоторое количество опытов, т.е. для нескольких комбинаций значений независимых переменных определить значения характеристики процесса v , которые при этом обозначают латинской буквой y , и подвергнуть эти данные соответствующей математической обработке. При этом возникает следующая трудность. В ходе эксперимента неизбежны флуктуации независимых переменных x_1, x_2, \dots, x_k и 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 прост-

ранстве (рис. 1.1). Возможные экспериментальные значения функции отклика $y=f(x_1, x_2)$ на рисунке не показаны, - при некоторых значениях x_1 и x_2 экспериментальные точки функции отклика y будут расположены выше гипотетической поверхности, при других - ниже. Поэтому по результатам выборки можно определить только выборочные коэффициенты уравнения регрессии, обозначаемые латинскими буквами $b_0, b_j, b_{1..m}, b_{jj}$:

$$\hat{y} = b_0 + \sum_{j=1}^k b_j x_j + \sum_{\substack{l=1 \\ m=2 \\ l=m}}^k b_{1..m} x_1 x_m + \sum_{j=1}^k b_{jj} x_j^2 \dots, \quad (1.4)$$

где \hat{y} - расчётное значение функции отклика. Выборочные коэффициенты $b_0, b_j, b_{1..m}, b_{jj}$ ещё называют **оценками** коэффициентов $\beta_0, \beta_j, \beta_{1..m}, \beta_{jj}$. Для определения генеральных коэффициентов $\beta_0, \beta_j, \beta_{1..m}, \beta_{jj}$ 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 v (см. рис. 1.1).

Рассмотрим подробнее особенность описания неизвестной функции $v=f(x_1, x_2, \dots, x_k)$ уравнением регрессии вида (1.4). Для примера рассмотрим некоторую гипотетическую зависимость $v=f(x_1, x_2)$, изображённую на рис. 1.1. Выделим область X возможных значений независимых переменных x_1 и x_2 , которым будет соответствовать поверхность F в системе координат v, x_1, x_2 . Выделим в пределах области X несколько участков поменьше, например a, b и c . Тогда значениям факторов x_1 и x_2 в пределах участка a будет соответствовать часть купола F_a , участку b будет соответствовать F_b , а c - F_c . Начиная исследование, экспериментатор, в общем случае, не имеет представления о форме поверхности функции v и о том, какое место его опыты занимают в пространстве, образованном независимыми переменными и функцией v . 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 X возможных

значений переменных 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 значительной степени зависят от масштаба измерения переменных и направления их варьирования. Другими словами, результат моделирования зависит от места расположения начала координат, масштаба и от направления осей в многофакторном пространстве. Эти (и другие) трудности явились стимулом поиска специальных методов планирования экспериментов.

Прежде чем переходить к подробному рассмотрению метода полного факторного эксперимента рассмотрим определения некоторых терминов. При первом чтении их можно пропустить.

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, $s^2_{ад}$:

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

$$s^2_{ад} = \sum_{i=1}^n (y_i - \hat{y}_i)^2 / (n - l), \quad (1.6)$$

где $n_{оп}$ - число опытов в выборке на *воспроизводимости*; 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 однород-

ны, математическая модель адекватна, если неоднородны - неадекватна. Более или менее объективным критерием однородности дисперсий является значение опытного критерия Фишера:

$$F_{\nu_1, \nu_2}^{\text{оп}} = \frac{S_{\text{ад}}^2}{S_{\text{оп}}^2}, \quad (1.7)$$

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. $S_{\text{ад}}^2$ и $S_{\text{оп}}^2$ однородны и, соответственно, модель адекватна эксперименту, если больше - нулевая гипотеза отклоняется, т.е. дисперсии неоднородны, модель неадекватна. См. также *Дисперсия адекватности*.

Величина параметрическая - общее название физических величин, приведённых к безразмерному виду путём того или иного соотношения с параметром. Параметрические величины получаются в результате нормирования переменных и приведения переменных. Если параметрические величины образуют 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, \dots, 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, \dots, 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 беконечность возможных анализов, опытов, измерений и т.п.

Дисперсия (< лат. disperse - рассеянно, разбросанно, там и сямя; 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

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

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

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

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 $\sigma^2 = DX$, 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

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

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 aequo - сравнивать, уравнивать, aequatus - приравненный, равный) - количественная характеристика расхождения экспериментальных значений функции отклика и значений функции отклика, рассчитанных по уравнению, параметры которого определены по выборке. См. также *Адекватность*.

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

$$s_{\text{оп}}^2 = \frac{1}{n_{\text{оп}} - 1} \sum_{i=1}^{n_{\text{оп}}} (y_i - \bar{y})^2; \quad (1.11)$$

$$\bar{y} = \frac{1}{n_{\text{оп}}} \sum_{i=1}^{n_{\text{оп}}} y_i, \quad (1.12)$$

где $n_{\text{оп}}$ - число опытов на воспроизводимость; $\nu_{\text{оп}} = n_{\text{оп}} - 1$ - 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

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

для непрерывной -

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

Для практического расчёта дисперсии воспроизводимости более удобна и точна формула

$$s_{0n}^2 = \frac{1}{n_{0n}(n_{0n}-1)} \left\{ n_{0n} \sum_{i=1}^{n_{0n}} y_i^2 - \left(\sum_{i=1}^{n_{0n}} y_i \right)^2 \right\}. \quad (1.15)$$

Для оценки достоверности результатов наблюдений одной дисперсии воспроизводимости недостаточно. Корень квадратный из дисперсии воспроизводимости называется *квадратичным отклонением* или *стандартным отклонением*. Начинаящие исследователи обычно с трудом развивают интуитивное восприятие численного значения дисперсии или стандартного отклонения. Является ли дисперсия воспроизводимости, равная, например, $77 \cdot 10^{-4}$? 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, $y_{ср}$. Если $|s_y| < y_{ср}$, то говорят о *значимом* отличии результатов наблюдений от нуля и *предварительную* оценку точности эксперимента осуществляют по отношению $s_y/y_{ср}$. Если это отношение лежит в пределах $3 \pm 4\%$, то в первом приближении результаты наблюдений считают воспроизводимыми; в противном случае необходимы дополнительные изыскания.

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

$$P = P\{ |\beta - b| < \epsilon_p \}, \quad (1.16)$$

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

$$\alpha = 1 - P = P\{ |\beta - b| > \epsilon_p \} \quad (1.17)$$

Иначе выражение (1.16) может быть интерпретировано как вероятность того, что истинное значение параметра β находится в пределах

$$b - \epsilon_p < \beta < b + \epsilon_p \quad (1.18)$$

где выборочный параметр b , по существу, *случайная величина*, а ошибка его определения $(\beta - b)$ в выражениях (1.16) и (1.17) также случайная величина. Интервал $I_p = b \pm \epsilon_p$ называется *доверительным интервалом*. Границы интервала $b_{\min} = b - \epsilon_p$ и $b_{\max} = b + \epsilon_p$ называются *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:

$$\frac{S_x}{\sqrt{n}} \cdot t_{\nu}^{\alpha} \quad (1.19)$$

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 b_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, который с вероятностью P "накрывает" неизвест-

ное 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):

$$P\{ |M_x - \bar{x}| < \epsilon_p \} = \int_{-\epsilon_p}^{+\epsilon_p} p(x) dx = P. \quad (1.20)$$

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

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

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:

$$\bar{x} - \frac{s_x}{\sqrt{n}} \cdot t < m_x < \bar{x} + \frac{s_x}{\sqrt{n}} \cdot t, \quad (1.21)$$

где t - критерий Стьюдента, случайная величина, зависящая только от

числа степеней свободы ν выборочной дисперсии и уровня значимости α . Из оценки доверительного интервала 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, Доверительный интервал, Квадратичное отклонение, Стандартное отклонение, Стьюдента критерий*.

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

$$\frac{(x_1-a)^2 + (x_2-a)^2 + \dots + (x_n-a)^2}{n} \quad (1.22)$$

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

$$x_{cp} = \bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n} \quad (1.23)$$

В теории вероятностей **квадратичное отклонение** σ_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 σ_x вероятность отклонений X от M_x , больших по абсолютной величине $k\sigma_x$, $k>0$, 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*.

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 статистiku и генетику Р. Фишеру, который в начале 20-х гг. обратил внимание на то, что диагональность матрицы $X^T X$ позволит получить некоррелированные коэффициенты b_j , а необходимым и достаточным условием диагональности матрицы $X^T X$ является ортогональность матрицы X . Другими словами, Р. Фишер предложил спланировать эксперимент таким образом, чтобы исходная матрица X была ортогональна. Развитие идеи планирования эксперимента привело к тому, что Г.Е.Р. Вох и К.В. 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

$$x_j^0 = \frac{x_j^{\max} + x_j^{\min}}{2}, \quad (1.24)$$

а интервал варьирования

$$\Delta x_j = \frac{x_j^{\max} - x_j^{\min}}{2}. \quad (1.25)$$

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

$$\Delta x_j \Rightarrow \Delta z_j = 1; \quad j=1, 2, \dots, k. \quad (1.26)$$

В факторном пространстве размерности $k+1$ точка с координатами $x_1^0, x_2^0, \dots, x_k^0$ называется центром плана. При переходе к кодированным переменным z_1, z_2, \dots, z_k , в эту точку без поворота перенос-

сится начало координат и в этой новой (кодированной) системе координат z_1, z_2, \dots, z_k все переменные изменяются в интервале от -1 до $+1$:

$$z_j = \frac{x_j - x_j^0}{\Delta x_j}; \quad (1.27)$$

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, \dots, 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, \dots . 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 жестко 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 .

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

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

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

$$M_x = \sum_{i=1}^{\infty} x_i p_i; \quad (1.28)$$

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

$$M_x = \int_{-\infty}^{\infty} xp(x) dx. \quad (1.29)$$

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. *Значимость коэффициента*.

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

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \cdot e^{-(M_x - x)^2 / 2\sigma^2} \quad (1.30)$$

с неизвестными параметрами M_x и σ^2 , то среднее арифметическое

$$\bar{x} = (x_1 + x_2 + \dots + x_n) / n \quad (1.31)$$

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

$$s_x^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 \quad (1.32)$$

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_x^2 уменьшить на единицу. Несмещённой оценкой для σ^2 служит

$$s_x^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2. \quad (1.33)$$

Нормализация - установление нормы, образца; приведение к норме, к нормальному состоянию, урегулирование; вид термической обработки стали; 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, Стандартизация случайной величины.

Нормирование переменных - частный случай преобразования физических величин к безразмерному виду путём деления исходной переменной величины на аналогичную постоянную. Локализация области варьирования фактора производится путём нормирования переменных, при этом минимальное значение фактора принимается за ноль, а максималь-

ное за единицу. В результате безразмерная нормированная величина изменяется в интервале от нуля до единицы:

$$z_j = \frac{x_j - x_j^{\min}}{x_j^{\max} - x_j^{\min}}. \quad (1.34)$$

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. Нормирование *случайных величин* производится путём деления их на *квадратичное отклонение*. См. также *Величина параметрическая, Нормализация, Приведение переменных, Стандартизация случайной величины*.

Опытная дисперсия - см. *Дисперсия воспроизводимости*.

Оценка - количественная характеристика *параметра*, получаемая по результатам *выборки*. Проблема оценки неизвестного параметра является одной из центральных в теории обработки результатов наблюдений. К оценкам параметров предъявляется комплекс требований. Важнейшие среди них *несмещённость, состоятельность и эффективность*.

Параметр (от греч. *παράμετρον* - 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. Параметры 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:

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

где в первом случае B - аналог теплоты конденсации, а во втором - энергия активации вязкого течения; коэффициент A - предэкспоненциальный множитель; коэффициент 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).

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*.

Причина (укр. причина, польск. przyczyna < чин, чинить. Лат. causa) – явление, вызывающее возникновение другого явления. Познавание *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 $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ 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 упрощенных, но правдоподобных допущениях: величина x является **контролируемой** величиной, значения которой заранее задаются при планировании эксперимента, а наблюдаемые значения Y представимы в виде

$$y_i = \varphi(x_i, \beta) + \varepsilon_i, \quad i=1, 2, \dots, n, \quad (1.35)$$

где величины ε_i характеризуют ошибки, независимые при различных измерениях и одинаково распределенные с нулевым средним и постоянной дисперсией σ^2 . В действительности, невозможно осуществить эксперимент, в котором отсутствовала бы ошибка измерения y 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 $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ представляют собой выборку из совокупности с некоторым двумерным распределением вероятностей. И в том, и в другом случае 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* недостаточен эффективен, целесообразно подобрать детерминистическую мо-

дель, отражающую физическую сущность процесса. Стандартный метод оценки линии регрессии основан на использовании полиномиальной модели ($n \geq 1$):

$$y = b_0 + b_1 x + b_2 x^2 + b_3 x^3 + \dots + b_n x^n. \quad (1.36)$$

Оценки b_0, b_1, \dots, b_n принято осуществлять методом наименьших квадратов. Многочлен (1.36) 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_1 значений $y_{1,1}, y_{1,2}, \dots, y_{1,n_1}$ величины y , то зависимость средних арифметических

$$\bar{y}_1 = (y_{1,1} + y_{1,2} + \dots + y_{1,n_1}) / n_1 \quad (1.37)$$

от x_1 является **регрессией** в статистическом понимании этого термина; в пределе $n_1=1$. 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 реально 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.

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. $X = \dots, 1973, 1974, 1975, \dots$, 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 слу-

чаях 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 характеристике. 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 $n \rightarrow \infty$ 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), (1.33), (1.39), (1.50) и т.п. См. также *Дисперсия воспроизводимости.*

Статистических гипотез, проверка (греч. *υποθεσις* - основание, принцип, предположение, гипотеза; *υποτιθησις* - 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 подходя-

щей 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_{b_j} , которое соответственно может быть и очень малым числом, и очень большим, вопрос в соизмеримости b_j и s_{b_j} . Поэтому, сохранив сущность, нулевую гипотезу сформулируем по другому и будем проверять гипотезу об отсутствии различия между коэффициентом b_j и его квадратичным отклонением s_{b_j} . Если статистику критерия построить в виде отношения

$$t_{0n} = \frac{|b_j|}{s_{b_j}}, \quad (1.38)$$

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 $\Theta = f(X_1, X_2, \dots, X_n)$ от результатов наблюдений X_1, X_2, \dots, 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. 1.2).

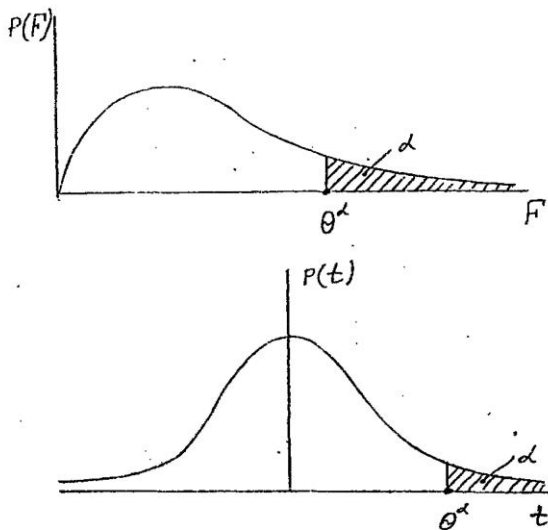


Рис. 1.2. Типичные F - и t -распределения. Заштрихована критическая область, составляющая α площади под кривой (общая площадь под кривой равна 1)

Область значений (x_1, x_2, \dots, x_n) , для которых $\theta(x_1, x_2, \dots, x_n) > \theta^\alpha$, 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, \dots . Для подтверждения этой гипотезы необходимо определить параметры математической модели, соответствую-

ющей природе изучаемого процесса, и проверить математическую модель 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, \dots, y_n и выборку размерности $n_{оп}$ для оценки точности эксперимента $y_1, y_2, \dots, y_{оп}$:

$$s^2_{оп} = \frac{1}{n_{оп}-1} \cdot \sum_{i=1}^{n_{оп}} (y_i - \bar{y})^2; \quad (1.39)$$

$$\bar{y} = \frac{1}{n_{оп}} \cdot \sum_{i=1}^{n_{оп}} y_i. \quad (1.40)$$

где $s^2_{оп}$ - несмещенная 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

$$F_{\nu_1, \nu_2}^{оп} = \frac{s^2_{ад}}{s^2_{оп}}; \quad (1.41)$$

$$s^2_{ад} = \frac{1}{n-1} \cdot \sum_{i=1}^n (y_i - \hat{y}_i)^2, \quad (1.42)$$

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); $y_{i, расч}$ - значения функции отклика рассчитанные по проверяемой модели с параметрами, определёнными по выборке y_1, y_2, \dots, y_n , причём каждому $y_{i, расч}$ соответствует $y_{i, эксп}$. Опытный критерий Фишера сравнивается с табличным значением, которое получается независимо с помощью формулы плотности распределения критерия Фишера:

$$p(F) = \frac{\Gamma\left(\frac{\nu_1 + \nu_2}{2}\right)}{\Gamma\left(\frac{\nu_1}{2}\right) \cdot \Gamma\left(\frac{\nu_2}{2}\right)} \cdot \frac{\nu_2^{2} \nu_1^{2} F^{\nu_1/2-1}}{(\nu_1 F + \nu_2)^{(\nu_1 + \nu_2)/2}}. \quad (1.43)$$

Если $F^{оп} < F^{\alpha}$, то мы оказываемся в области принятия нулевой гипотезы, т.е. математическая модель с уровнем значимости α адекватна; если

$F^{0n} > F^\alpha$, то мы попадаем в критическую область, область отклонения нулевой гипотезы - модель неадекватна. Если $F^{0n} < F^\alpha$, 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, \dots, x_n - взаимно независимые нормально распределённые случайные величины с неизвестными параметрами M_x и σ^2 , т.е. $N(M_x, \sigma^2)$. 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 :

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i; \quad (1.44)$$

$$s_x^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2, \quad (1.45)$$

где x_{cp} - оценка математического ожидания M_x , а s_x^2 - оценка генеральной дисперсии σ^2 . В тех случаях, когда $s_x \approx |x_{cp}|$ или $s_x > |x_{cp}|$, 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:

$$\bar{x} - \frac{s_x}{\sqrt{n}} \cdot t \leq M_x \leq \bar{x} + \frac{s_x}{\sqrt{n}} \cdot t, \quad (1.46)$$

где t - случайная величина, зависящая только от числа степеней свободы ν выборочной дисперсии и уровня значимости α . Формулой (1.46) пользуются практически при обработке экспериментальных данных для определения *доверительного интервала* x_{cp} . Очевидно, что выражение

$$\frac{s_x}{\sqrt{n}} \cdot t_{\nu}^{\alpha}, \quad (1.47)$$

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

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, \dots, X_n и если требуется по этим данным проверить гипотезу H_0 , согласно которой совместное распределение величин X_1, X_2, \dots, X_n обладает некоторым определённым свойством, то соответствующий статистический критерий **конструируется** с помощью подходящим образом подобранной функции $\theta = f(X_1, X_2, \dots, X_n)$. 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 заклю-

чается 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 $\alpha=0,05$. В теории статистической проверки гипотез уровень значимости называется *вероятностью ошибки первого рода*. Вероятность такой ошибки не больше **принятого** уровня значимости. Например, при $\alpha=0,05$ 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 ($\alpha=0,01$) или даже в одном случае из двадцати ($\alpha=0,05$), 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 ($\alpha > 0,1$), 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. 1.2, но в общем случае он может быть другим. Предста-

вим 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 $\alpha=0,05$ (рис. 1.3, 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.

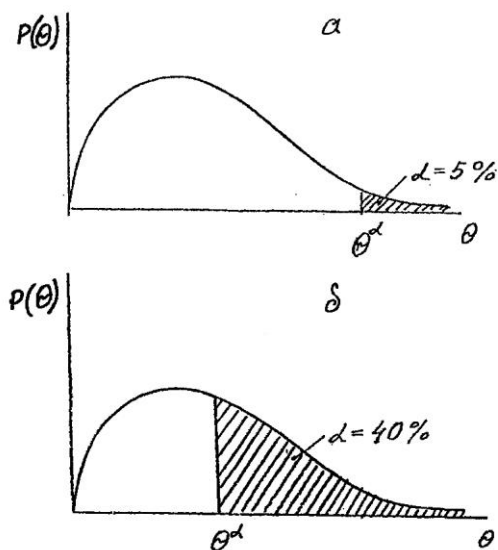


Рис. 1.3. Распределение статистики гипотетического критерия θ с критической областью α отклонения гипотезы о бесперспективности бурения

Если принять уровень значимости побольше, например $\alpha=0,40$ (рис. 1.3, 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 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 $\alpha=0,9$ [20].

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. 1.4). 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 записывают, например, так: для левой критической области $\theta^{1-\alpha/2}$, для правой - $\theta^{\alpha/2}$ [2].

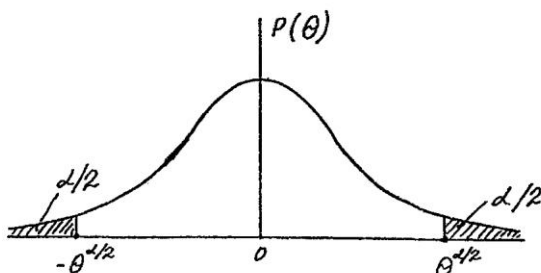


Рис. 1.4. Левая $-\theta^{\alpha/2}$ и правая $\theta^{\alpha/2}$ критические области гипотетического двустороннего критерия θ^{α}

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) – движущая сила, причина какого-либо процесса, явления; существенное обстоятельство в каком-либо процессе, явлении; независимая переменная физическая величина или *аргумент*. При наличии трёх и более аргументов принято говорить о пространстве независимых переменных или *факторном пространстве*.

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.

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 $y=f(x_1, x_2, \dots, x_k)$, 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, \dots, y_m для каждой конкретной точки фак-

торного пространства $(x_1, x_2, \dots, x_k)_1$ при экспериментальном определении, где k - количество независимых переменных (факторов) или размерность факторного пространства. При этом y_1, y_2, \dots, y_m рассматриваются как случайные величины, имеющие то или иное распределение вероятностей. Если в результате обработки экспериментальных данных получено уравнение регрессии $y=f(x_1, x_2, \dots, x_k)$, то y также называют функцией отклика, поскольку между функцией отклика $y=f(x_1, x_2, \dots, x_k)$ и пространством $Y=f(X)$ 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, Уровень значимости, Статистические гипотезы проверки*).

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 ν , а число параметров, определённых по выборке - латинской буквой l ; таким образом, $\nu=n-l$. Число параметров, определённых по выборке, ещё называют *числом связей*, наложенных на выборку. Так вот, с точки зрения теории информации число степеней свободы равно числу параметров, которое ещё можно определить по выборке после той или иной обработки, а с точки зрения математической статистики ν 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 ν . Если, например, математическое ожи-

данные M_x оценивается по результатам пяти независимых наблюдений

$$\bar{x} = (x_1 + x_2 + x_3 + x_4 + x_5) / 5, \quad (1.48)$$

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:

$$x_5 - \bar{x} = 5\bar{x} - (x_1 + x_2 + x_3 + x_4). \quad (1.49)$$

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 зависящая от рассматриваемой выборки (например m_x), т.е. оценка определяется независимым путём. В таких случаях для выборочной дисперсии следует пользоваться формулой (1.32):

$$s_x^2 = \frac{1}{n} \sum_{i=1}^n (x_i - m_x)^2; \quad (1.50)$$

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.

1.1. Обоснование метода наименьших квадратов

С помощью метода наименьших квадратов осуществляется восстановление зависимости, полученной в результате экспериментальных исследований, с целью описания её аналитической функцией. В случае парной зависимости $y=f(x)$ это может быть либо прямая линия, либо кривая (естественно, с большим или меньшим "разбросом" экспериментальных точек относительно прямой или кривой). Если ограничиться небольшим участком зависимости $y=f(x)$, то вполне возможно, что нелинейную зависимость, с достаточной для практических целей точностью, можно будет описать уравнением прямой линии $y=b_0+b_1x$ (в пределах ошибки определения x_1).

В случае зависимости вида $y=f(x_1, x_2)$ можно построить рисунок (хотя это и сложнее, чем в случае парной зависимости) и также определиться в возможности представления участка экспериментальной зависимости $y=f(x_1, x_2)$ плоскостью, т.е. уравнением $y=b_0+b_1x_1+b_2x_2$. На этом возможности человека в представлении пространства на бумаге и в уме заканчиваются. Уже в случае с тремя факторами исследователю приходится действовать вслепую (в смысле визуализации зависимости). Поэтому на начальных этапах исследований целесообразно ограничиваются небольшим участком поверхности, для которого может быть справедливо линейное приближение. Соответственно, для описания этой поверхности можно использовать уравнение *линейной множественной регрессии*:

$$y=b_0+b_1x_1+b_2x_2+\dots+b_kx_k, \quad (1.51)$$

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

$$y=b_0+b_1x_1+b_2x_2+\dots+b_kx_k+b_{1,2}x_1x_2+b_{2,3}x_2x_3+\dots+b_{k-1,k}x_{k-1}x_k+b_{1,1}x_1^2+b_{2,2}x_2^2+\dots+b_{k,k}x_k^2, \quad (1.52)$$

где $b_0, b_1, \dots, b_k, b_{1,2}, b_{2,3}, \dots, b_{k-1,k}, b_{1,1}, b_{2,2}, \dots, b_{k,k}$ - искомые коэффициенты уравнения регрессии, а, точнее, оценки соответствующих генеральных коэффициентов $\beta_j, \beta_{1,m}$ 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 $y=f(x_1, x_2, \dots, x_k)$. Эта модель *линейна относительно параметров* β_j , но не обязана быть линейной относительно x_1, x_2, \dots, x_k . Дело в том, что, по существу, метод наименьших квадратов даёт несмещённые оценки параметров, являющихся линейными функциями от наблюдений и имеющих минимальную дисперсию. Метод НК - случай так называемой линейной модели, когда наблюдения являются линейными функциями неизвестных параметров β_j . Другими словами, существенна линейность только относительно параметров, тогда как возможно более "естественная" линейность функции отклика y от x_1, x_2, \dots, 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, \dots, x_k . Например непосредственно полиномиальная зависимость

$$y=b_0+b_1x+b_2x^2+b_3x^3+\dots+b_nx^n \quad (1.53)$$

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

$$y=b_0+b_1x_1+b_2x_2+\dots+b_kx_k+b_{12}x_1x_2+b_{23}x_2x_3+\dots+b_{k-1,k}x_{k-1}x_k+ \\ +b_{11}x_1^2+b_{22}x_2^2+\dots+b_{kk}x_k^2 \quad (1.54)$$

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

$$y=b_0+b_1x_1+b_1^2x_2+b_2x_3 \quad (1.55)$$

с точки зрения метода НК линейным не является, поскольку в него входят и β_1 и β_1^2 . Аналогично, модель нелинейная по своей сущности:

$$y=b_0+b_1x_1+b_2x_1^2+b_3\sin x_1\cos x_2 \quad (1.56)$$

с точки зрения метода НК является линейной. Причина кажущегося противоречия ("модель линейная относительно β_j , нелинейная относительно x_j " и наоборот) кроется в изначальной постановке задачи оценки неизвестных параметров распределения методом НК. Эта задача решается путём дифференцирования некоторой функции $\Phi=f(\beta_j, x_1, y_1)$ по оцен-

кам параметров b_j искомого уравнения, причём предполагается, что каждая оценка b_j в функции Φ встречается один раз и только в первой степени.

Подробнее. Дело 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 $y=f(x_1, x_2, \dots, x_k)$ 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, которые различаются по точности и, что немаловажно, по трудоёмкости. Наибольшее распространение получил метод наименьших квадратов (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).

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

$$\Phi = \sum_{i=1}^n \left(y_i - f(x_1, x_2, \dots, x_k) \right)_i^2, \quad (1.57)$$

где в рассматриваемом случае

$$f(x_1, x_2, \dots, x_k)_i = \sum_{j=0}^k b_j x_{i,j}. \quad (1.58)$$

Как общий принцип он формулируется следующим образом: в качестве оценки вектора неизвестных параметров b_0, b_1, \dots, b_k в некотором выражении $\varphi(\mathbf{X}, \mathbf{Y}, \mathbf{B})=0$, где \mathbf{X} и \mathbf{Y} - результаты наблюдений, а \mathbf{B} - вектор параметров, следует взять такое $\mathbf{B}_{расч}$, которое обращает в минимум сумму (1.57). 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)

$$M_x = \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i, \quad (1.59)$$

а оценкой генеральной дисперсии является выборочная дисперсия

$$s^2_x = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2. \quad (1.60)$$

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

$$\Phi = \sum_{i=1}^n \left(y_i - bx_i \right)^2 = \min. \quad (1.61)$$

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:

$$\Phi = \sum_{i=1}^n \left(y_i - (b_0 + b_1 x_i) \right)^2 = \min. \quad (1.62)$$

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 $y=f(x_1, x_2, \dots, x_k)$, тем или иным сечением. Поэтому следующим уравнением, описывающим некоторые процессы в природе, является уравнение линейной множественной регрессии

$$y = b_0 + b_1 x_1 + b_2 x_2 + b_3 x_3 + \dots + b_k x_k. \quad (1.63)$$

Человеку, живущему в четырёхмерном пространстве-времени и привыкшего строить соответствующие мысленные модели, трудно создавать и анализировать многомерную мысленную модель, образ многомерного пространства. Человеку, может быть, не проще, но эффективнее анализировать отдельные зависимости $y=f(x_1)$, $y=f(x_2)$, ..., $y=f(x_k)$, обобщать и делать выводы впоследствии. Именно поэтому основной прогресс в развитии методов поиска параметров уравнения вида (1.63) 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 (1.63) можно представить так

$$Y = BX + \epsilon, \quad (1.64)$$

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

$$M(\epsilon) = 0 \quad (1.65)$$

и матрицей рассеяния

$$V(\epsilon) = M(\epsilon\epsilon^T) = \sigma^2 I, \quad (1.66)$$

где \mathbf{I} - единичная матрица размерности $n \times n$, а σ^2 - дисперсия ошибок наблюдений. Увеличение числа столбцов матрицы \mathbf{X} на единицу обусловлено введением в неё первого, а, точнее, нулевого столбца, состоящего из единиц, что вызвано необходимостью вычисления свободного члена b_0 . Условия (1.65) и (1.66) соответствуют предположению о том, что ε_i некоррелированы (т.е. взаимонезависимы), имеют нулевые средние и одинаковую дисперсию σ^2 . Раньше, в 50-е - 60-е гг., считалось, что на элементы матрицы \mathbf{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 $\pm 0.1^\circ$, эксперименты проводят с интервалом $5 \cdot 10^\circ$, иногда больше.

Метод НК состоит в минимизации скалярной суммы квадратов

$$\Phi = \sum_{i=1}^n \left(y_i - \sum_{j=0}^k b_j x_{i,j} \right)^2 = (\mathbf{Y} - \mathbf{B}\mathbf{X})^T (\mathbf{Y} - \mathbf{B}\mathbf{X}) \quad (1.67)$$

по компонентам вектора \mathbf{B} . Поскольку функция $\Phi \geq 0$ при любых b_0, b_1, \dots, b_k , у неё должен быть хотя бы один минимум. Условием минимума является условие $\Phi / \mathbf{B} = 0$.

Прежде, чем перейти к детальному рассмотрению условия минимума функции Φ , целесообразно обсудить некоторые моменты относительно искомых параметров уравнения b_0, b_1, \dots, b_k и переменных величин x_1 и y_1 . В процессе дифференцирования условия (1.67) параметры b_j будущего уравнения рассматриваются как переменные величины, функции результатов наблюдений, а измеряемые в процессе эксперимента случайные величины x_1 и y_1 рассматриваются как постоянные. Выполняя диф-

ференцирование, получим

$$2\mathbf{X}^T(\mathbf{Y}-\mathbf{X}\mathbf{B})=0, \quad (1.68)$$

откуда находим вектор-столбец оценок \mathbf{B} :

$$\mathbf{B}=(\mathbf{X}^T\mathbf{X})^{-1}(\mathbf{X}^T\mathbf{Y}). \quad (1.69)$$

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

2. ПОЛНЫЙ ФАКТОРНЫЙ ЭКСПЕРИМЕНТ

Анализ формул пассивного эксперимента [2, 11, 17] показывает, что при вычислении каждого коэффициента b_j используются элементы всех строк и всех столбцов матрицы X . Это означает, что при вычислении каждого j -того коэффициента кроме значений j -того фактора используются значения и всех остальных факторов, никакого отношения к j -тому коэффициенту не имеющие, т.е. коэффициенты уравнения регрессии не могут быть определены независимо друг от друга или, другими словами, все коэффициенты взаимосвязаны, коррелированы. Следствием этого является то, что и доверительные границы для каждого j -того коэффициента в отдельности можно установить только после фиксации всех остальных коэффициентов. Если один или более коэффициентов уравнения регрессии оказывается незначимым, то просто так его удалить нельзя, нужно обратиться к исходной матрице X , удалить из неё j -тый столбец, соответствующий незначимому коэффициенту b_j , и повторить весь расчёт. До появления вычислительных машин необходимость пересчёта была достаточно существенным недостатком.

Обобщая, основные недостатки пассивного эксперимента можно свести к корреляции коэффициентов, отсутствию ротатабельности (т.е. зависимость коэффициентов от поворота системы координат в факторном пространстве) и к невозможности при сохранении статистических оценок исключить опыт или удалить фактор x_j [17, с.60].

2.1. Кодирование переменных

В начале 20-х гг. английский генетик и статистик Р. Фишер обратил внимание на то, что диагональность матрицы $X^T X$ позволит получить некоррелированные коэффициенты b_j , а необходимым и достаточным условием диагональности матрицы $X^T X$ является ортогональность матрицы X . Другими словами, Р. Фишер предложил спланировать эксперимент таким образом, чтобы исходная матрица X была ортогональна. Развитие идеи планирования эксперимента привело к тому, что G.E.P. Box и K.B. Wilson предложили все физические переменные представлять в виде +1 и -1, т.е. предложили так называемое кодирование переменных [1, 2, 12]. 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, а минимальным значениям соответственно -1. Например, если исследуется влияние температуры, давления и концентрации реагента на какой-либо процесс в интервале 60÷120°С, 120÷150 атм и 1,0÷1,5 моль/л, то значениям температуры 120°С, давления 150 атм и концентрации реагента 1,5 моль/л будет соответствовать одно и то же кодированное значение +1 (так называемый *верхний уровень*), а значениям температуры 60°С, 120 атм и 1,0 моль/л будет соответствовать -1 (так называемый *нижний уровень*). Соответственно, *нулевому уровню* будут соответствовать значения 90°С, 135 атм и 1,25 моль/л. В общем случае *центр плана* по каждому фактору

$$x_j^0 = \frac{x_j^{\max} + x_j^{\min}}{2}; \quad (2.1)$$

а интервал варьирования:

$$\Delta x_j = \frac{x_j^{\max} - x_j^{\min}}{2}; \quad (2.2)$$

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

$$\Delta x_j = \Delta z_j = 1; \quad j=1, 2, \dots, k. \quad (2.3)$$

В факторном пространстве размерности $k+1$ точка с координатами $x_1^0, x_2^0, \dots, x_k^0$ называется центром плана. При переходе к кодированным переменным z_1, z_2, \dots, z_k , в эту точку без поворота переносится начало координат, и в этой новой (кодированной) системе координат z_1, z_2, \dots, z_k все переменные изменяются в интервале от -1 до +1:

$$z_j = \frac{x_j - x_j^0}{\Delta x_j}. \quad (2.4)$$

Исключение составляет только функция отклика y , которая измеряется и участвует в последующих расчётах в прежних физических единицах.

На вопрос необходимого количества опытов, наверное, естественным ответом будет перебор всех возможных сочетаний факторов на верхнем и нижнем уровнях, т.е. общее число опытов

$$n=2^k, \quad (2.5)$$

где k - количество независимых переменных. Очевидно, что в нашем примере общее число опытов будет равно $n=2^3=8$.

2.2. Построение плана экспериментов

Попробуем теперь матрицу

$$X = \begin{pmatrix} x_{1,0} & x_{1,1} & x_{1,2} & x_{1,3} \\ x_{2,0} & x_{2,1} & x_{2,2} & x_{2,3} \\ x_{3,0} & x_{3,1} & x_{3,2} & x_{3,3} \\ x_{4,0} & x_{4,1} & x_{4,2} & x_{4,3} \\ x_{5,0} & x_{5,1} & x_{5,2} & x_{5,3} \\ x_{6,0} & x_{6,1} & x_{6,2} & x_{6,3} \\ x_{7,0} & x_{7,1} & x_{7,2} & x_{7,3} \\ x_{8,0} & x_{8,1} & x_{8,2} & x_{8,3} \end{pmatrix} \quad (2.6)$$

представить в виде матрицы кодированных переменных Z . Как и в случае пассивного эксперимента, нулевой столбец (столбец фиктивной переменной) должен состоять из +1, а первый, второй и третий столбцы должны состоять из +1 и -1. Попробуем взять за основу ортогональную матрицу

$$Z = \begin{pmatrix} z_0 & z_1 \\ +1 & +1 \\ +1 & -1 \end{pmatrix} \Leftrightarrow Y = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}, \quad (2.7)$$

где первому столбцу будет соответствовать фиктивная переменная z_0 , а второму - переменная z_1 . Развивая аналогию соответствия строк матрицы Z условиям экспериментов, поставим в соответствие первой строке значение функции отклика y_1 , а второй строке - y_2 . При таком подходе в декартовой системе координат y - z матрицам Z и Y (2.7) будет соответствовать прямая линия с координатами $(+1, y_1)$, $(-1, y_2)$, если иметь в виду, что кодированное значение $z_{1,1}=+1$, а $z_{2,1}=-1$ (рис. 2.1).

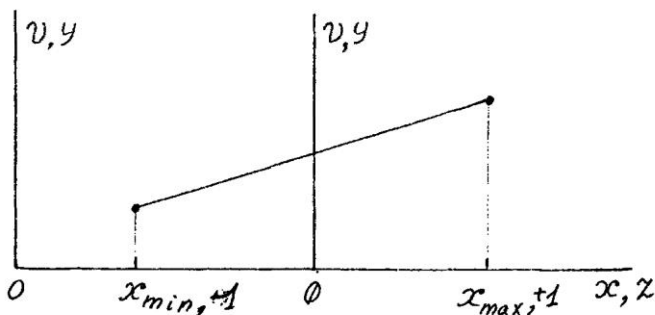


Рис. 2.1. Гипотетический пример представления
однофакторного плана
в декартовой системе координат $y=f(x)$;
 U - гипотетическое значение функции $U=f(x)$

Если эту матрицу повторить дважды и добавить третий столбец со значениями $z_2=+1, +1, -1, -1$

$$Z = \begin{matrix} & z_0 & z_1 & z_2 \\ \begin{matrix} + \\ + \\ + \\ + \end{matrix} & \begin{vmatrix} +1 & +1 & +1 \\ +1 & -1 & +1 \\ +1 & +1 & -1 \\ +1 & -1 & -1 \end{vmatrix} & \Leftrightarrow Y = \begin{vmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{vmatrix}, & (2.8) \end{matrix}$$

то полученная матрица Z также будет ортогональной (есть и другие приёмы построения матрицы плана ПФЭ [2, 11]). В системе координат $y-z_1, z_2$ матрице (2.8) и значениям y_1, y_2, y_3, y_4 в общем случае будет соответствовать некоторая поверхность отклика Y , а в частном случае - плоскость. В плоскости z_1-z_2 столбцам z_1 и z_2 матрицы (2.8) будет соответствовать квадрат (в кодированных значениях или в безразмерном масштабе), в центре которого находится начало координат (рис. 2.2). Не следует смущаться отсутствием строгости в рассуждениях "факторное пространство кодированных переменных" и то же пространство с физическими переменными y_1, y_2, \dots, y_n . В данном случае это не более, как методический приём.

Повторяя матрицу (2.8) дважды и, аналогично, добавляя четвертый столбец, получим (см. след. стр., матрица (2.9))

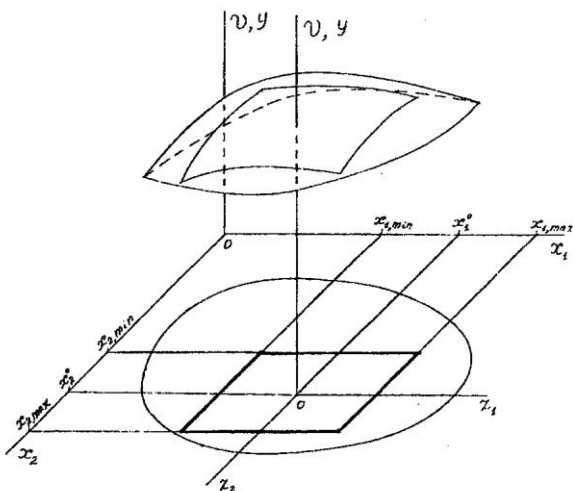
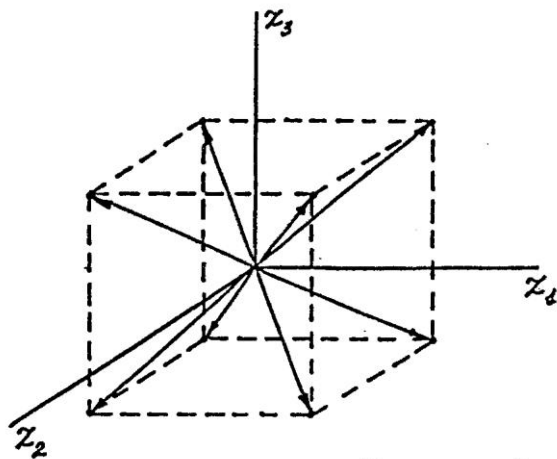


Рис. 2.2. Пример построения двухфакторного плана в системе координат $y=f(x_1, x_2)$; v - гипотетическое значение функции $v=f(x_1, x_2)$

$$Z = \begin{vmatrix} Z_0 & Z_1 & Z_2 & Z_3 \\ +1 & +1 & +1 & +1 \\ +1 & -1 & +1 & +1 \\ +1 & +1 & -1 & +1 \\ +1 & -1 & -1 & +1 \\ +1 & +1 & +1 & -1 \\ +1 & -1 & +1 & -1 \\ +1 & +1 & -1 & -1 \\ +1 & -1 & -1 & -1 \end{vmatrix} \Leftrightarrow Y = \begin{vmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \\ y_6 \\ y_7 \\ y_8 \end{vmatrix} \quad (2.9)$$

Геометрически матрица (2.9) в кодированном трёхмерном факторном пространстве представляет собой куб (рис. 2.3): очевидно, что в привычном для человека представлении трёхмерного пространства изобразить на рисунке поверхность функции отклика y невозможно. По аналогии можно построить матрицу кодированных значений Z_1, Z_2, \dots, Z_k для любого числа независимых переменных. Геометрической интерпретацией такой матрицы будет правильный многогранник, имеющий 2^k вершин, в центре которого будет находиться начало кодированной системы координат. Очевидно, что строки такой матрицы соответствуют условиям отдельных экспериментов, а число строк равно числу возможных сочетаний значений k факторов на верхнем и нижнем уровнях. Матрица,

Рис. 2.3. Пример трёхфакторного плана $Y=f(Z_1, Z_2, Z_3)$

подобная (2.8), (2.9) и т.д., называется матрицей плана *полного факторного эксперимента* (ПФЭ). Отметим основные свойства такой матрицы.

$$\sum_{i=1}^n z_{i,l} z_{i,m} = 0; \quad l \neq m; \quad l, m = 0, 1, 2, \dots, k; \quad (2.10)$$

$$\sum_{i=1}^n z_{i,j} = 0; \quad j \neq 0; \quad j = 1, 2, \dots, k; \quad (2.11)$$

$$\sum_{i=1}^n z_{i,j}^2 = n; \quad j = 0, 1, 2, \dots, k. \quad (2.12)$$

Свойство (2.10) - равенство нулю скалярных произведений всех столбцов матрицы Z , - свойство ортогональности, является главным свойством матрицы плана эксперимента. Это важнейшая характеристика плана, поскольку при этом матрица $Z^T Z$ становится диагональной, что, в свою очередь, приводит к независимости коэффициентов уравнения регрессии друг от друга и к одинаковой ошибке в их определении. Второе свойство (2.11) - условие симметричного расположения всех независимых переменных относительно центра эксперимента. Наконец, третье свойство (2.12) - равенство сумм квадратов элементов для всех столбцов. Следствием третьего условия является равенство всех диагональных элементов матрицы $Z^T Z$ числу наблюдений n и равенство

всех диагональных элементов матрицы $(Z^T Z)^{-1} a_{jj} = 1/n$.

В заключение этого раздела важно заметить следующее. Перед осуществление экспериментов по плану (2.7), (2.9) и т.д. необходимо произвести процедуру *рандомизации* (от англ. *random* - сделанный или выбранный наугад, случайный, беспорядочный). Дело в том, что условия проведения экспериментов по планам (2.7), (2.9) и т.д. изменяются строго закономерно; при реализации эксперимента эта закономерность в неявном виде войдет в окончательное уравнение регрессии как дополнительный, но неучтенный фактор. Для того, чтобы этот дополнительный, а, по своей сути, несуществующий фактор, перевести в область случайных ошибок, закономерность чередования условий экспериментов необходимо нарушить и выполнять эксперименты в случайном порядке. Рандомизацию можно произвести с помощью таблицы случайных чисел или элементарного жребия.

2.3. Вычисление коэффициентов по плану ПФЭ

Рассмотрим методику вычисления коэффициентов на примере двухфакторного плана.

Результаты экспериментов, реализованных по плану матрицы (2.8) можно представить в виде матрицы-столбца

$$Y = \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \end{pmatrix}. \quad (2.13)$$

Искомый вектор-столбец коэффициентов уравнения регрессии согласно (2.3):

$$B = (Z^T Z)^{-1} (Z^T Y) = \begin{pmatrix} b_0 \\ b_1 \\ b_2 \end{pmatrix}. \quad (2.14)$$

Преобразуем матрицу условий экспериментов Z и матрицу-столбец результатов Y согласно (2.14):

$$Z^T = \begin{pmatrix} +1 & +1 & +1 & +1 \\ +1 & -1 & +1 & -1 \\ +1 & +1 & -1 & -1 \end{pmatrix}; \quad Z^T Z = \begin{pmatrix} +4 & 0 & 0 \\ 0 & +4 & 0 \\ 0 & 0 & +4 \end{pmatrix};$$

$$(\mathbf{Z}^T \mathbf{Z})^{-1} = \begin{vmatrix} 1/4 & 0 & 0 \\ 0 & 1/4 & 0 \\ 0 & 0 & 1/4 \end{vmatrix}; \quad \mathbf{Z}^T \mathbf{Y} = \begin{vmatrix} (+y_1 + y_2 + y_3 + y_4) \\ (+y_1 - y_2 + y_3 - y_4) \\ (+y_1 + y_2 - y_3 - y_4) \end{vmatrix};$$

$$\mathbf{B} = \begin{vmatrix} 1/4 \cdot (+y_1 + y_2 + y_3 + y_4) \\ 1/4 \cdot (+y_1 - y_2 + y_3 - y_4) \\ 1/4 \cdot (+y_1 + y_2 - y_3 - y_4) \end{vmatrix}.$$

В общем виде j -тый коэффициент уравнения регрессии вычисляется скалярным произведением j -того столбца матрицы \mathbf{Z} на столбец наблюдений \mathbf{Y} :

$$b_j = \frac{1}{n} \sum_{t=1}^n z_{1,j} y_t \quad (2.15)$$

Очевидно, что b_j некоррелированы. Рассуждая аналогично (3.6) и (3.7) [17], получим:

$$s_{b_j}^2 = \begin{vmatrix} 1/n & 0 & \dots & 0 & \dots & 0 \\ 0 & 1/n & \dots & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 1/n & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 0 & \dots & 1/n \end{vmatrix} \cdot s_y^2 = \frac{1}{n} s_{оп}^2. \quad (2.16)$$

Свойство (2.16) называется свойством *равноошибочности*. Оно в данном случае означает, что все коэффициенты уравнения регрессии определяются с одинаковой степенью точности. Диагональность матрицы $\mathbf{Z}^T \mathbf{Z}$ означает, что все ковариации равны нулю, т.е. все коэффициенты определяются независимо друг от друга, и незначимые коэффициенты удаляются из уравнения без пересчёта оставшихся.

2.4. Проверка коэффициентов уравнения на значимость

Проверка коэффициентов уравнения регрессии на значимость позволяет выявить те коэффициенты b_j , которые по тем или иным причинам незначимо отличаются от нуля. Независимо от абсолютного значения коэффициента незначимое отличие от нуля будет соответствовать факту нахождения нуля в *доверительном интервале* коэффициента b_j .

$$b_j - s_{b_j} \cdot t_{\nu_{оп}}^{\alpha} < \beta_j < b_j + s_{b_j} \cdot t_{\nu_{оп}}^{\alpha}, \quad (2.17)$$

где β_j - генеральный j -тый коэффициент; s_{b_j} - стандартное отклонение коэффициента b_j ; t^{α} - критерий Стьюдента для уровня значимости α и числа степеней свободы $\nu_{оп}$ (сравните с соотношением 1.46). В соответствии с соотношением (2.17) подходящей нулевой гипотезой будет гипотеза об отсутствии различия между нулём и коэффициентом b_j - $H_0: b_j=0$. Но практически, нулевую гипотезу формулируют в виде отсутствия различия между коэффициентом b_j и его стандартным отклонением s_{b_j} - $H_0: b_j=s_{b_j}$.

Регрессионный анализ полученного уравнения начинается с вычисления дисперсии коэффициентов $s_{b_j}^2$, характеризующей ошибку определения любого коэффициента. В отличие от пассивного эксперимента матрица $(Z^T Z)^{-1}$ диагональна и все ковариации коэффициентов равны нулю, т.е. все коэффициенты независимы друг от друга. Согласно (2.16):

$$s_{b_j}^2 = \frac{1}{n} \cdot s_{оп}^2.$$

где $s_{оп}^2$ - дисперсия воспроизводимости, вычисляемая по формулам (1.11) и (1.12)

$$s_{оп}^2 = \frac{1}{n_{оп}-1} \sum_{i=1}^{n_{оп}} (y_i - \bar{y})^2; \quad \bar{y} = \frac{1}{n_{оп}} \sum_{i=1}^{n_{оп}} y_i.$$

где $n_{оп}$ - число опытов на воспроизводимость, а $n_{оп}-1 = \nu_{оп}$ - число степеней свободы дисперсии воспроизводимости. Дисперсию воспроизводимости ещё называют опытной дисперсией (от Experiment - опыт, термина, принятого в англо-американской литературе), но значение русского слова "воспроизводимость" больше соответствует физической сущности дисперсии воспроизводимости. Если при пассивном эксперименте нет принципиального значения, при каком сочетании значений факторов ставить опыты на воспроизводимость, то при использовании методов планирования эксперимента эти опыты следует ставить в центре плана. При такой стратегии исследования будет проще перейти к композиционным планам Д. Бокса и К. Уилсона. Кроме этого, при насыщении плана (см. ниже) можно будет проверить нуль-гипотезу об адекватности уравнения путём сравнения свободного члена b_0 со средним значением

выборки в центре плана и, наконец, можно будет ещё проверить нуль-гипотезу о том, что сумма всех коэффициентов регрессии $\sum \beta_{j,j}$ при квадратичных членах z^2_j равна нулю. Этот способ оценки нелинейности основан на том, что свободный член b_0 , вычисляемый по формуле (2.15)

$$b_0 = \frac{1}{n} \sum_{i=1}^n z_{i,0} y_i - \frac{1}{n} \sum_{i=1}^n y_i \quad (2.18)$$

является оценкой, смешанной с квадратичными эффектами:

$$b_0 = \beta_0 + \sum_{j=1}^k \beta_{j,j}, \quad (2.19)$$

поскольку столбцы z^2_j неотличимы от столбца z_0 . Поэтому разность

$$b_0 - \bar{y}_0 = \sum_{j=1}^k \beta_{j,j} \quad (2.20)$$

в какой-то степени может служить мерой кривизны поверхности [11].

Собственно проверка коэффициента на значимость заключается в сравнении опытного критерия Стьюдента с табличным:

$$t_{оп} = \frac{|b_j|}{s_{b_j}} \Leftrightarrow t_{\nu_{оп}}^{\alpha}, \quad (2.21)$$

где табличное значение критерия Стьюдента берётся для числа степеней свободы дисперсии воспроизводимости $\nu_{оп}$ и уровня значимости α . Нулевая гипотеза формулируется как отсутствие различия между коэффициентом b_j и его квадратичным отклонением s_{b_j} :

$$H_0: b_j = s_{b_j}.$$

Естественной альтернативной гипотезой в этом случае будет неравенство:

$$H_1: b_j \neq s_{b_j}.$$

Если опытный критерий Стьюдента меньше табличного, то у нас нет оснований отклонить основную проверяемую гипотезу, т.е. коэффициент b_j незначимо отличается от нуля. Если опытный критерий Стьюдента больше табличного, то он попадает в критическую область, в область отклонения нулевой гипотезы, и мы вынуждены принять альтернативную

гипотезу $b_j \neq 0$ (см. рис. 1.2 и 1.4). С доверительной вероятностью $P=1-\alpha$ коэффициент b_j значимо отличается от нуля. Проверку можно осуществить и с помощью доверительного отклонения

$$S_{b_j}^{\alpha} = S_{b_j} \cdot t_{\nu_{оп}}^{\alpha}, \quad (2.22)$$

что более наглядно. Коэффициент b_j незначим, если

$$|b_j| < S_{b_j}^{\alpha}. \quad (2.23)$$

Оценку значимости коэффициентов можно производить и с помощью критерия Фишера:

$$F_{\nu_1, \nu_2}^{оп. j} = \frac{S_j^2}{S_{оп}^2}, \quad (2.24)$$

где $S_j^2 = n b_j^2 / \nu_1$ - дисперсия, определяющая вклад, вносимый j -тым коэффициентом уравнения регрессии в суммарную дисперсию, создаваемую всеми коэффициентами. Число степеней свободы ν_2 равно числу степеней свободы дисперсии воспроизводимости, а $\nu_1 = 1$. Дело в том, что в соответствии с общими принципами дисперсионного анализа сумма квадратов

$$n \cdot \sum_{j=0}^k b_j^2$$

с числом степеней свободы $\nu = k+1$ в данном случае раскладывается на $k+1$ составляющих. Подобное разложение возможно, поскольку все коэффициенты некоррелированы [11].

2.5. Проверка уравнения регрессии на адекватность

Проверка уравнения регрессии на адекватность необходима для уверенности в том, что расчёты по полученному уравнению будут в той или иной степени соответствовать действительности. Сущность вопроса, как придать большей объективности "той или иной степени соответствия", достаточно подробно рассмотрена в разделах "Статистических гипотез проверка" и "Уровень значимости", поэтому перейдём непосредственно к проверке уравнения регрессии на адекватность. Пос-

ледняя, по существу, сводится к проверке гипотезы об однородности дисперсии воспроизводимости $s^2_{оп}$ и дисперсии адекватности $s^2_{ад}$.

Дисперсия адекватности $s^2_{ад}$ вычисляется по формуле (1.6)

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

где n - число опытов в выборке (т.е. число опытов поставленных с целью вычисления коэффициентов уравнения регрессии); $\nu_{ад}$ - число степеней свободы дисперсии адекватности. В общем случае число степеней свободы $\nu = n - l$, где l - число связей, наложенных на выборку (общее число параметров определённых по выборке); в частном случае $\nu = n - k - 1$, где k - количество независимых переменных. Расчётное значение функции отклика вычисляется по формуле

$$\hat{y}_1 = b_0 + \sum_{j=1}^k b_j z_{1,j} + \sum_{\substack{l=1 \\ m=2 \\ l \neq m}}^k b_{l,m} z_{1,l} z_{1,m}, \quad (2.26)$$

где $z_{1,j}$ - кодированные значения факторов согласно матрице независимых переменных (2.8), (2.9) и т.п. для каждого опыта, от $l=1$ до $l=n$.

В нашем случае имеются две выборки - экспериментальная и рассчитанная:

$$y_1, y_2, \dots, y_n;$$

$$\hat{y}_1, \hat{y}_2, \dots, \hat{y}_n.$$

Необходимо проверить, являются ли эти две выборки частями одной и той же совокупности или это части разных совокупностей. Другими словами, если ошибка в определении оценок коэффициентов полученного уравнения настолько велика, что рассчитанные значения функции отклика формируют другую совокупность, отличную от первой, экспериментально исследуемой, то и некоторые параметры этих выборок должны отличаться. Одним из параметров, характеризующих специфические свойства выборки, является выборочная дисперсия. Дисперсии выборок из одной и той же совокупности называются однородными, из разных выборок - неоднородными. Если дисперсии $s^2_{ад}$ и $s^2_{оп}$ окажутся однородными, то это будет означать, что расчёт по найденному уравнению

соответствует экспериментальным данным, т.е. уравнение адекватно, если неоднородны - уравнение неадекватно и пользоваться им нельзя.

Проверяется нулевая гипотеза H_0 о равенстве генеральных дисперсий σ^2_1 и σ^2_2 :

$$H_0: \sigma^2_1 = \sigma^2_2.$$

Для того, чтобы подтвердить эту гипотезу, необходимо доказать однородность дисперсий $s^2_{ад}$ и $s^2_{оп}$, а для того, чтобы отвергнуть - доказать значимость различия $s^2_{ад}$ и $s^2_{оп}$. Доверительная вероятность $P=1-\alpha$ имеет непосредственное отношение к сущности вопроса "той или иной степени соответствия" расчётов по полученному уравнению реальной действительности. Величиной доверительной вероятности P следует задаваться до проверки гипотезы, например 0,9, 0,95, 0,99 и т.д. В условиях нулевой гипотезы $\sigma^2_1 = \sigma^2_2$, $\sigma^2_1/\sigma^2_2 = 1$ и $F = (s^2_1/s^2_2)$, т.е. F -распределение может быть непосредственно использовано для оценки отношения выборочных дисперсий s^2_1/s^2_2 :

$$F_{\nu_{ад}, \nu_{оп}}^{оп} = \frac{s^2_{ад}}{s^2_{оп}}, \quad (2.27)$$

Опытное значение критерия Фишера $F^{оп}$ сравнивается с табличным значением F^α для чисел степеней свободы $\nu_{ад}$ и $\nu_{оп}$ и уровня значимости α , который применительно к данному случаю характеризует предельное соотношение однородных дисперсий. Если опытное значение критерия Фишера меньше табличного, то дисперсии $s^2_{ад}$ и $s^2_{оп}$ являются однородными и соответственно гипотеза H_0 принимается. Применительно к нашему случаю это означает, что уравнение регрессии адекватно. Если $F^{оп} > F^\alpha$, т.е. опытный критерий Фишера попадает в критическую область (см. рис. 1.2 и 1.3), то дисперсии $s^2_{ад}$ и $s^2_{оп}$ считаются неоднородными и гипотеза H_0 отклоняется, т.е. полученное уравнение регрессии неадекватно. В этом случае возможны три выхода: попробовать другой метод обработки данных, предположить другой вид функциональной зависимости $y=f(x_1, x_2, \dots, x_k)$, повысить точность экспериментов или продолжить эксперименты в других точках факторного пространства.

Следует обратить внимание на тот факт, что уравнение регрессии, полученное по методу полного факторного эксперимента, может оказаться адекватным и для нелинейной функции отклика в том случае, если эксперимент грубый, если велика дисперсия воспроизводимости $s^2_{оп}$.

В заключение раздела о проверке уравнения на адекватность следует обратить внимание на тот факт, что в соотношении (2.27) в числителе - дисперсия адекватности $s^2_{ад}$, а в знаменателе - дисперсия воспроизводимости $s^2_{оп}$. Запись вида (2.27) общепринята потому, что практически дисперсия адекватности бывает почти всегда больше дисперсии воспроизводимости. А вот практическое превышение дисперсии адекватности над дисперсией воспроизводимости объясняется тем, что дисперсия адекватности включает в себя как ошибки эксперимента, так и ошибки, обусловленные приближенным характером уравнения регрессии. Дисперсия воспроизводимости $s^2_{оп}$ включает в себя только ошибки эксперимента. Поскольку таблицы F -распределения содержат значения F_{ν_1, ν_2} превосходящие 1, то для значений $F^{оп}$ меньших единицы следует использовать соотношение [21]

$$F_{\nu_1, \nu_2}^{\alpha} = 1 / F_{\nu_2, \nu_1}^{1-\alpha} \quad (2.28)$$

2.6. Декодирование уравнения регрессии

Уравнение, полученное по методу ПФЭ (или методом планирования первого порядка по Д. Боксу и К. Уилсоу), в кодированных переменных имеет вид

$$y = b_0 + b_1 z_1 + b_2 z_2 + \dots + b_k z_k \quad (2.29)$$

где переменные z_j при обработке результатов могут принимать только фиксированные значения -1 или $+1$ согласно матриц (2.7), (2.8), (2.9) и т.д. После проверки коэффициентов уравнения (2.29) на значимость и проверки уравнения на адекватность необходимо вернуться в исходную, физическую систему координат с помощью соотношения (2.4):

$$y = b_0 + b_1 \left(\frac{x_1 - x_1^0}{\Delta x_1} \right) + b_2 \left(\frac{x_2 - x_2^0}{\Delta x_2} \right) + \dots + b_k \left(\frac{x_k - x_k^0}{\Delta x_k} \right) = \quad (2.30)$$

$$= \left(b_0 - b_1 \frac{x_1^0}{\Delta x_1} - b_2 \frac{x_2^0}{\Delta x_2} - \dots - b_k \frac{x_k^0}{\Delta x_k} \right) + \frac{b_1}{\Delta x_1} x_1 + \frac{b_2}{\Delta x_2} x_2 + \dots + \frac{b_k}{\Delta x_k} x_k$$

2.7. Расширение полного факторного эксперимента

У рассмотренного плана имеется ряд существенных недостатков, в том числе линейность уравнения относительно факторов x_j и значительное число степеней свободы ν дисперсии адекватности:

$$S_{ад}^2 = \frac{1}{n-k-1} \sum_{i=1}^n (y_i - \hat{y}_i)^2. \quad (2.31)$$

Так, при $k=2$ - $\nu=n-k-1=1$, при $k=3$ - $\nu=n-k-1=4$, при $k=4$ - $\nu=n-k-1=11$. Очевидно, что при больших k в выборке остаётся неиспользованным значительный информационный потенциал. Последний недостаток носит название *ненасыщенности* ПФЭ, в основном этот недостаток преодолевается с помощью *дробных реплик* (см. ниже).

Линейность уравнения регрессии относительно факторов X_j отчасти преодолевается путём введения в него парных и тройных произведений факторов, так называемых *эффектов парного взаимодействия* $b_{1,2}$, $b_{2,3}$, $b_{1,3}, \dots$ и *эффектов тройного взаимодействия* $b_{1,2,3}$, $b_{2,3,4}, \dots$. Так, для $k=3$ будем иметь

$$y = b_0 + b_1 Z_1 + b_2 Z_2 + b_3 Z_3 + b_{12} Z_1 Z_2 + b_{23} Z_2 Z_3 + b_{1,3} Z_1 Z_3 + b_{1,2,3} Z_1 Z_2 Z_3. \quad (2.32)$$

Для вычисления эффектов взаимодействия следует расширить матрицу плана (2.9) путём введения в неё соответствующих произведений элементов столбцов.

$$Z = \begin{matrix} & Z_0 & Z_1 & Z_2 & Z_3 & Z_1 Z_2 & Z_2 Z_3 & Z_1 Z_3 & Z_1 Z_2 Z_3 \\ \left. \begin{array}{l} +1 \\ +1 \\ +1 \\ +1 \\ +1 \\ +1 \\ +1 \\ +1 \end{array} \right\} & \begin{array}{l} +1 \\ +1 \\ +1 \\ -1 \\ +1 \\ -1 \\ +1 \\ -1 \end{array} & \begin{array}{l} +1 \\ -1 \\ +1 \\ -1 \\ +1 \\ -1 \\ +1 \\ -1 \end{array} & \begin{array}{l} +1 \\ +1 \\ -1 \\ +1 \\ -1 \\ -1 \\ -1 \\ +1 \end{array} & \begin{array}{l} +1 \\ -1 \\ -1 \\ +1 \\ -1 \\ -1 \\ +1 \\ +1 \end{array} & \begin{array}{l} +1 \\ +1 \\ -1 \\ +1 \\ -1 \\ -1 \\ +1 \\ +1 \end{array} & \begin{array}{l} +1 \\ -1 \\ +1 \\ -1 \\ -1 \\ +1 \\ -1 \\ +1 \end{array} & \begin{array}{l} +1 \\ -1 \\ +1 \\ -1 \\ -1 \\ +1 \\ -1 \\ +1 \end{array} & \begin{array}{l} +1 \\ +1 \\ +1 \\ -1 \\ -1 \\ -1 \\ -1 \\ -1 \end{array} \end{matrix}. \quad (2.33)$$

Нетрудно убедиться, что расширенная матрица ПФЭ (2.33) ортогональна, т.е. все коэффициенты определяются независимо друг от друга и с одинаковой ошибкой. Каждый j -тый коэффициент уравнения регрессии вычисляется аналогично (2.15), а каждый коэффициент $b_{1,m}$ при

парном произведении $z_l z_m$ - скалярным произведением столбцов l и m матрицы Z на столбец наблюдений Y :

$$b_{l,m} = \frac{1}{n} \sum_{i=1}^n z_{l,i} z_{m,i} y_i. \quad (2.34)$$

Проверка на значимость коэффициентов $b_{l,m}$ при парных произведениях осуществляется по формулам (2.16), (2.21)+(2.24).

Особенности будут только с проверкой уравнения (2.32) на адекватность, поскольку в случае значимости всех коэффициентов не остаётся ни одной степени свободы - знаменатель в формуле (2.25) равен нулю: $\nu = n - l = 0$, где l - число связей, накладываемых на выборку (число параметров, определенных по выборке), в данном случае 8. Диалектика научного познания проявляется здесь в том, что, определив по выборке из 8 опытов 8 параметров, т.е. изъяв из выборки весь информационный потенциал, исследователь лишается возможности проверить достоверность полученных результатов. В таком случае нет другого выхода, как поставить ещё один опыт, на этот раз в центре плана, и сравнить его результат y_0 со свободным членом b_0 :

$$t_{\text{эксн}} = \frac{|b_0 - y_0|}{S_{\text{он}}} \sqrt{n} < t_{\nu, \alpha}^{\text{он}}. \quad (2.35)$$

Если экспериментальное значение критерия Стьюдента меньше табличного, взятого для числа степеней свободы дисперсии воспроизводимости, то полученное уравнение адекватно эксперименту с уровнем значимости α (приимается нулевая гипотеза), если больше - неадекватно. Практически, эффекты тройного (и выше) взаимодействия факторов редко оказываются значимыми, обычно их и не вычисляют.

Строго говоря, одного опыта в центре плана явно недостаточно, оптимально все опыты на воспроизводимость ставить в центре плана. При этом можно проверить нуль-гипотезу о равенстве нулю суммы всех коэффициентов $\sum b_{j,j}$ при квадратичных эффектах X^2_j . Дело в том, что свободный член уравнения регрессии b_0 , вычисляемый по формуле (2.15), является смешанной оценкой: $b_0 = \beta_0 + \sum \beta_{j,j}$, так как элементы столбца z_0 не отличимы от элементов столбцов z^2_j . Поэтому разность:

$$b_0 - \bar{y}_0 \rightarrow \sum \beta_{j,j} \quad (2.36)$$

в некоторой степени является мерой кривизны поверхности отклика.

Достаточным основанием для отклонения гипотезы линейности функции отклика является факт значительного отличия от нуля любого коэффициента $b_{1,m}$ при парном произведении факторов $z_1 z_m$.

Декодирование уравнения, полученного по расширенному плану Д. Бокса и К. Уилсона, производится аналогично (2.30) с помощью соотношения (2.4). Необходимо отметить одну, не отмеченную в литературе, особенность декодирования уравнений типа (2.32): бывают случаи незначимости коэффициента b_j и значимости, например, $b_{j,m}$. Согласно требованиям регрессионного анализа, незначимый коэффициент b_j необходимо исключить, а значимый коэффициент $b_{j,m}$ оставить, но при декодировании уравнения коэффициент при j -том факторе появится, и j -тый фактор как бы возвращается в уравнении регрессии. Вопрос о действительной значимости j -того фактора остаётся открытым.

3. ДРОБНЫЙ ФАКТОРНЫЙ ЭКСПЕРИМЕНТ

Рассмотрение идеи кодирования переменных по Д. Боксу и К. Уил-сону естественно приводит к выводу, что перебор всех возможных сочетаний факторов на верхнем и нижнем уровнях соответствует общему числу опытов $n=2^k$, где k - количество независимых переменных. Очевидно, что с увеличением числа факторов общее число опытов нарастает в геометрической прогрессии и, соответственно, возрастает число степеней свободы дисперсии адекватности. Другими словами, возрастает неиспользованный информационный потенциал выборки, что в некотором смысле невыгодно (см. табл. 3.1). Конечно, чем больше массив наблюдений, тем точнее определяются искомые параметры, но практически это всего лишь оценки истинных значений, а не сами значения. А поскольку математические ожидания можно определить только реализовав совокупность, то следует экспериментировать в реальном оптимуме. Дело в том, что значительное превышение числа опытов над минимально необходимым $n=k+2$ или $n=l+1$ (для проверки уравнения на адекватность нужна хотя бы одна степень свободы) обусловлено перебором всех возможных сочетаний факторов на верхнем и нижнем уровнях. Вопрос в том, какие именно опыты из матрицы, например (2.9), можно исключить без существенной потери точности. Очевидно, что не все комбинации опытов, оставляемых в матрице плана, будут равноценны по точности получаемых в результате коэффициентов. Очевидно также, что сокращение числа опытов следует производить не механически, а по некоторому закону. Было бы хорошо, чтобы этот закон также позволял предвидеть последствия тех или иных комбинаций исключаемых опытов. В 1945 г. D.J. Finnty предложил снизить число экспериментов путём использования так называемых *дробных реплик* и предложил закон формирования матрицы плана [15]. Таким законом является *генерирующее соотношение* вида, например, $Z_k = Z_1 Z_m$ или $Z_k = Z_1 Z_m Z_u$. Это значит, что при построении матрицы трёхфакторного плана

Таблица 3.1.

Зависимость количества опытов и числа степеней свободы от размерности факторного пространства

k	n	ν
2	4	1
3	8	4
4	16	11
5	32	26
6	64	57
...
k	2^k	$n-k-1$

Следует обратить внимание на тот факт, что в качестве закона построения столбца для четвертого фактора можно использовать ещё три генерирующих соотношения: $Z_4 = Z_1 Z_2$, $Z_4 = Z_1 Z_3$ и $Z_4 = Z_2 Z_3$. Очевидно, что четыре возможных генерирующих соотношения неравноценны. Настало время обсудить вопрос о недостатках метода дробного факторного эксперимента (ДФЭ).

Первая потеря очевидна - дробная реплика на основе того или иного генерирующего соотношения лишает возможности ввести в уравнение соответствующее парное произведение или, другими словами, проследить эффект парного взаимодействия, поскольку соответствующие столбцы идентичны. Следствием является невозможность описания нелинейной поверхности или приходится ограничиваться небольшим участком поверхности, чтобы можно было получить линейное приближение.

Теперь обсудим вопрос о предвидении последствий тех или иных комбинаций исключаемых опытов. Если для плана (3.2) генерирующее соотношение $Z_3 = Z_1 Z_2$ умножить на Z_3 , то

$$Z_3^2 = Z_1 Z_2 Z_3, \quad (3.4)$$

где $Z_3^2 = 1$. Выражение $1 = Z_1 Z_2 Z_3$ называется *определяющим контрастом*, с его помощью удобно определять *смещение оценок*, умножая на соответствующий фактор:

$$\begin{aligned} Z_1 = Z_1^2 Z_2 Z_3 &\Rightarrow Z_1 = Z_2 Z_3 \Rightarrow b_1 = \beta_1 + \beta_{23}; \\ Z_2 = Z_1 Z_2^2 Z_3 &\Rightarrow Z_2 = Z_1 Z_3 \Rightarrow b_2 = \beta_2 + \beta_{13}; \\ Z_3 = Z_1 Z_2 Z_3^2 &\Rightarrow Z_3 = Z_1 Z_2 \Rightarrow b_3 = \beta_3 + \beta_{12}. \end{aligned} \quad (3.5)$$

Очевидно также, что

$$1 = Z_1 Z_2 Z_3 \Rightarrow b_0 = \beta_0 + \beta_{123}. \quad (3.6)$$

Дело в том, что для полуреплики (3.2) столбец для произведения $Z_2 Z_3$ равен столбцу Z_1 , столбец $Z_1 Z_3$ равен столбцу Z_2 , а столбец $Z_1 Z_2$ равен столбцу Z_3 . Термин "смещение оценок" означает, что, например, оценка b_1 неизвестного параметра β_1 включает в себя три составляющих: собственно параметр β_1 , параметр β_{23} при произведении факторов $Z_2 Z_3$ и ошибку ϵ_{b_1} (3.5). Не касаясь вопроса о том, в каких долях в оценку b_1 входят β_1 , β_{23} и ошибка ϵ_{b_1} , следует констатировать тот факт, что в реальных задачах эффекты парных взаимодействий редко равны нулю, и смещение оценок типа (3.5) нежелательно.

Аналогично, для плана (3.3) и генерирующего соотношения $Z_4 = Z_1 Z_2 Z_3$ имеем

$$\begin{aligned}
 Z_4^2 &= Z_1 Z_2 Z_3 Z_4 \Rightarrow Z_0 = Z_1 Z_2 Z_3 Z_4 \Rightarrow b_0 = \beta_0 + \beta_{1234}; \\
 Z_1 &= Z_1^2 Z_2 Z_3 Z_4 \Rightarrow Z_1 = Z_2 Z_3 Z_4 \Rightarrow b_1 = \beta_1 + \beta_{234}; \\
 Z_2 &= Z_1 Z_2^2 Z_3 Z_4 \Rightarrow Z_2 = Z_1 Z_3 Z_4 \Rightarrow b_2 = \beta_2 + \beta_{134}; \\
 Z_3 &= Z_1 Z_2 Z_3^2 Z_4 \Rightarrow Z_3 = Z_1 Z_2 Z_4 \Rightarrow b_3 = \beta_3 + \beta_{124}; \\
 Z_4 &= Z_1 Z_2 Z_3 Z_4^2 \Rightarrow Z_4 = Z_1 Z_2 Z_3 \Rightarrow b_4 = \beta_4 + \beta_{123}; \\
 Z_1 Z_2 &= Z_1^2 Z_2^2 Z_3 Z_4 \Rightarrow Z_1 Z_2 = Z_3 Z_4 \Rightarrow b_{12} = \beta_{12} + \beta_{34}; \\
 Z_1 Z_3 &= Z_1^2 Z_2 Z_3^2 Z_4 \Rightarrow Z_1 Z_3 = Z_2 Z_4 \Rightarrow b_{13} = \beta_{13} + \beta_{24}; \\
 Z_2 Z_3 &= Z_1 Z_2^2 Z_3^2 Z_4 \Rightarrow Z_2 Z_3 = Z_1 Z_4 \Rightarrow b_{23} = \beta_{23} + \beta_{14}.
 \end{aligned} \tag{3.7}$$

Для генерирующих соотношений $Z_4 = Z_1 Z_2$, $Z_4 = Z_1 Z_3$ и $Z_4 = Z_2 Z_3$ смешения оценок будут следующими:

при $Z_4 = Z_1 Z_2$

$$\begin{aligned}
 Z_4^2 &= Z_1 Z_2 Z_4 \Rightarrow 1 = Z_1 Z_2 Z_4 \Rightarrow b_0 = \beta_0 + \beta_{124}; \\
 Z_1 &= Z_1^2 Z_2 Z_4 \Rightarrow Z_1 = Z_2 Z_4 \Rightarrow b_1 = \beta_1 + \beta_{24}; \\
 Z_2 &= Z_1 Z_2^2 Z_4 \Rightarrow Z_2 = Z_1 Z_4 \Rightarrow b_2 = \beta_2 + \beta_{14}; \\
 Z_3 &= Z_1 Z_2 Z_3 Z_4 \Rightarrow b_3 = \beta_3 + \beta_{1234}; \\
 Z_4 &= Z_1 Z_2 Z_4^2 \Rightarrow Z_4 = Z_1 Z_2 \Rightarrow b_4 = \beta_4 + \beta_{12}; \\
 Z_1 Z_2 &= Z_1^2 Z_2^2 Z_4 \Rightarrow Z_1 Z_2 = Z_4 \Rightarrow b_{12} = \beta_{12} + \beta_4; \\
 Z_1 Z_3 &= Z_1^2 Z_2 Z_3 Z_4 \Rightarrow Z_1 Z_3 = Z_2 Z_3 Z_4 \Rightarrow b_{13} = \beta_{13} + \beta_{234}; \\
 Z_1 Z_4 &= Z_1^2 Z_2 Z_4^2 \Rightarrow Z_1 Z_4 = Z_2 \Rightarrow b_{14} = \beta_{14} + \beta_2; \\
 Z_2 Z_3 &= Z_1 Z_2^2 Z_3 Z_4 \Rightarrow Z_2 Z_3 = Z_1 Z_3 Z_4 \Rightarrow b_{23} = \beta_{23} + \beta_{134}; \\
 Z_2 Z_4 &= Z_1 Z_2^2 Z_4^2 \Rightarrow Z_2 Z_4 = Z_1 \Rightarrow b_{24} = \beta_{24} + \beta_1; \\
 Z_3 Z_4 &= Z_1 Z_2 Z_3 Z_4^2 \Rightarrow Z_3 Z_4 = Z_1 Z_2 Z_3 \Rightarrow b_{34} = \beta_{34} + \beta_{123}.
 \end{aligned} \tag{3.8}$$

при $Z_4 = Z_1 Z_3$

$$\begin{aligned}
 Z_4^2 &= Z_1 Z_3 Z_4 \Rightarrow 1 = Z_1 Z_3 Z_4 \Rightarrow b_0 = \beta_0 + \beta_{134}; \\
 Z_1 &= Z_1^2 Z_3 Z_4 \Rightarrow Z_1 = Z_3 Z_4 \Rightarrow b_1 = \beta_1 + \beta_{34}; \\
 Z_2 &= Z_1 Z_2 Z_3 Z_4 \Rightarrow b_2 = \beta_2 + \beta_{1234}; \\
 Z_3 &= Z_1 Z_3^2 Z_4 \Rightarrow Z_3 = Z_1 Z_4 \Rightarrow b_3 = \beta_3 + \beta_{14};
 \end{aligned} \tag{3.9}$$

$$\begin{aligned}
 Z_4 &= Z_1 Z_3 Z_4^2 \Rightarrow Z_4 = Z_1 Z_3 \Rightarrow b_4 = \beta_4 + \beta_{13}; \\
 Z_1 Z_2 &= Z_1^2 Z_2 Z_3 Z_4 \Rightarrow Z_1 Z_2 = Z_2 Z_3 Z_4 \Rightarrow b_{12} = \beta_{12} + \beta_{234}; \\
 Z_1 Z_3 &= Z_1^2 Z_3^2 Z_4 \Rightarrow Z_1 Z_3 = Z_4 \Rightarrow b_{13} = \beta_{13} + \beta_4; \\
 Z_1 Z_4 &= Z_1^2 Z_3 Z_4^2 \Rightarrow Z_1 Z_4 = Z_3 \Rightarrow b_{14} = \beta_{14} + \beta_3; \\
 Z_2 Z_3 &= Z_1 Z_2 Z_3^2 Z_4 \Rightarrow Z_2 Z_3 = Z_1 Z_2 Z_4 \Rightarrow b_{23} = \beta_{23} + \beta_{124}; \\
 Z_2 Z_4 &= Z_1 Z_2 Z_3 Z_4^2 \Rightarrow Z_2 Z_4 = Z_1 Z_2 Z_3 \Rightarrow b_{24} = \beta_{24} + \beta_{123}; \\
 Z_3 Z_4 &= Z_1 Z_3^2 Z_4^2 \Rightarrow Z_3 Z_4 = Z_1 \Rightarrow b_{34} = \beta_{34} + \beta_1;
 \end{aligned} \tag{3.9}$$

при $Z_4 = Z_2 Z_3$

$$\begin{aligned}
 Z_4^2 &= Z_2 Z_3 Z_4 \Rightarrow 1 = Z_2 Z_3 Z_4 \Rightarrow b_0 = \beta_0 + \beta_{234}; \\
 Z_1 &= Z_1 Z_2 Z_3 Z_4 \Rightarrow Z_1 = Z_1 Z_2 Z_3 Z_4 \Rightarrow b_1 = \beta_1 + \beta_{1234}; \\
 Z_2 &= Z_2^2 Z_3 Z_4 \Rightarrow Z_2 = Z_3 Z_4 \Rightarrow b_2 = \beta_2 + \beta_{34}; \\
 Z_3 &= Z_2 Z_3^2 Z_4 \Rightarrow b_3 = \beta_3 + \beta_{24}; \\
 Z_4 &= Z_2 Z_3 Z_4^2 \Rightarrow Z_4 = Z_2 Z_3 \Rightarrow b_4 = \beta_4 + \beta_{23}; \\
 Z_1 Z_2 &= Z_1^2 Z_2 Z_3 Z_4 \Rightarrow Z_1 Z_2 = Z_1 Z_3 Z_4 \Rightarrow b_{12} = \beta_{12} + \beta_{134}; \\
 Z_1 Z_3 &= Z_1 Z_2 Z_3^2 Z_4 \Rightarrow Z_1 Z_3 = Z_1 Z_2 Z_4 \Rightarrow b_{13} = \beta_{13} + \beta_{124}; \\
 Z_1 Z_4 &= Z_1 Z_2 Z_3 Z_4^2 \Rightarrow Z_1 Z_4 = Z_1 Z_2 Z_3 \Rightarrow b_{14} = \beta_{14} + \beta_{123}; \\
 Z_2 Z_3 &= Z_2^2 Z_3^2 Z_4 \Rightarrow Z_2 Z_3 = Z_4 \Rightarrow b_{23} = \beta_{23} + \beta_4; \\
 Z_2 Z_4 &= Z_2^2 Z_3 Z_4^2 \Rightarrow Z_2 Z_4 = Z_3 \Rightarrow b_{24} = \beta_{24} + \beta_3; \\
 Z_3 Z_4 &= Z_2 Z_3^2 Z_4^2 \Rightarrow Z_3 Z_4 = Z_2 \Rightarrow b_{34} = \beta_{34} + \beta_2.
 \end{aligned} \tag{3.10}$$

В реальных задачах коэффициенты при тройных произведениях факторов оказываются незначимыми значительно чаще, чем коэффициенты при парных произведениях. Анализ выражений (3.9), (3.10) показывает, что генерирующее соотношение $Z_4 = Z_1 Z_2 Z_3$ позволяет получить несмешанные оценки для линейных эффектов и смешанные для парных. Для остальных трёх генерирующих соотношений свойственно то, что оценки линейных эффектов факторов, участвующих в генерирующем соотношении, обязательно смешаны с соответствующими эффектами парных взаимодействий. В лучшем положении оказываются оценки параметров при факторах, не участвующих в генерирующем соотношении: эти оценки и оценки эффектов парного взаимодействия фактора с остальными смешаны только

с эффектами тройного и более взаимодействий. Например, если важно выделить влияние всех линейных эффектов, то следует использовать генерирующее соотношение $Z_4 = Z_1 Z_2 Z_3$, а если важно выделить влияние фактора, например, Z_2 и его парных произведений с остальными, то в качестве генерирующего соотношения следует принять $Z_4 = Z_1 Z_3$.

Очевидно, что прежде чем планировать эксперимент с использованием дробных реплик типа 2^{4-1} , необходимо каким-либо образом оценить влияние всех факторов на процесс и проанализировать последствия выбора того или иного генерирующего соотношения. Значительно лучше обстоит дело с дробными репликами при $k > 5$. В полурепликах с генерирующим соотношением $Z_5 = Z_1 Z_2 Z_3 Z_4$ все линейные эффекты и эффекты парного взаимодействия смешаны только с эффектами тройного и более высокого порядков:

$$2^{5-1} \Rightarrow Z_5 = Z_1 Z_2 Z_3 Z_4 \Rightarrow 1 = Z_1 Z_2 Z_3 Z_4 Z_5;$$

$$Z_0 = Z_1 Z_2 Z_3 Z_4 Z_5 \Rightarrow b_0 = \beta_0 + \beta_{12345}; \quad Z_1 = Z_2 Z_3 Z_4 Z_5 \Rightarrow b_1 = \beta_1 + \beta_{2345};$$

$$Z_2 = Z_1 Z_3 Z_4 Z_5 \Rightarrow b_2 = \beta_2 + \beta_{1345}; \quad Z_3 = Z_1 Z_2 Z_4 Z_5 \Rightarrow b_3 = \beta_3 + \beta_{1245};$$

$$Z_4 = Z_1 Z_2 Z_3 Z_5 \Rightarrow b_4 = \beta_4 + \beta_{1235}; \quad Z_5 = Z_1 Z_2 Z_3 Z_4 \Rightarrow b_5 = \beta_5 + \beta_{1234};$$

$$Z_1 Z_2 = Z_3 Z_4 Z_5 \Rightarrow b_{12} = \beta_{12} + \beta_{345}; \quad Z_1 Z_3 = Z_2 Z_4 Z_5 \Rightarrow b_{13} = \beta_{13} + \beta_{245};$$

$$Z_1 Z_4 = Z_2 Z_3 Z_5 \Rightarrow b_{14} = \beta_{14} + \beta_{235}; \quad Z_1 Z_5 = Z_2 Z_3 Z_4 \Rightarrow b_{15} = \beta_{15} + \beta_{234};$$

$$Z_2 Z_3 = Z_1 Z_4 Z_5 \Rightarrow b_{23} = \beta_{23} + \beta_{145}; \quad Z_2 Z_4 = Z_1 Z_3 Z_5 \Rightarrow b_{24} = \beta_{24} + \beta_{135};$$

$$Z_2 Z_5 = Z_1 Z_3 Z_4 \Rightarrow b_{25} = \beta_{25} + \beta_{134}; \quad Z_3 Z_4 = Z_1 Z_2 Z_5 \Rightarrow b_{34} = \beta_{34} + \beta_{125};$$

$$Z_3 Z_5 = Z_1 Z_2 Z_4 \Rightarrow b_{35} = \beta_{35} + \beta_{124}; \quad Z_4 Z_5 = Z_1 Z_2 Z_3 \Rightarrow b_{45} = \beta_{45} + \beta_{123}.$$

В подобных случаях возможно использование четвертьреплик. Более подробно см. [2, 11].

4. ПРИМЕР РАСЧЁТА

В научной лаборатории исследовалось влияние концентрации реагента, температуры и давления в процессе затвердевания цементного раствора на свойства цементного камня методом полного факторного эксперимента (ПФЭ). Один из планов ПФЭ осуществлялся при следующих пределах изменения переменных:

- концентрация реагента $c_p = 1,0 \pm 1,5\%$ мас.;
- температура $t = 60 \pm 120^\circ\text{C}$;
- давление $p = 120 \pm 150$ ат.

При отладке методики эксперимента в центре плана поставлено 5 опытов с результатами: 0,527; 0,528; 0,523; 0,522; 0,525. При этом достигнута точность эксперимента, соответствующая дисперсии воспроизводимости $s^2_{\text{оп}} = 0,65 \cdot 10^{-5}$ с числом степеней свободы $\nu = 4$.

Требуется получить уравнение регрессии вида

$$y = b_0 + b_1 c_p + b_2 t + b_3 p + b_{1,2} c_p t + b_{1,3} c_p p + b_{2,3} t p.$$

Введём обозначения:

- $x_1 = c_p, \%$;
- $x_2 = t, ^\circ\text{C}$;
- $x_3 = p, \text{ ат.}$

4.1. Кодирование переменных

Координаты центра плана вычисляются по формуле (2.1):

$$x_1^0 = \frac{x_1^{\max} + x_1^{\min}}{2} = \frac{1,0 + 1,5}{2} = 1,25\%;$$

$$x_2^0 = \frac{x_2^{\max} + x_2^{\min}}{2} = \frac{60 + 120}{2} = 90^\circ\text{C};$$

$$x_3^0 = \frac{x_3^{\max} + x_3^{\min}}{2} = \frac{120 + 150}{2} = 135 \text{ ат.},$$

а интервалы варьирования - формуле (2.2):

$$\Delta x_1 = \frac{x_1^{\max} - x_1^{\min}}{2} = \frac{1,5 - 1,0}{2} = 0,25\%;$$

$$\Delta x_2 = \frac{x_2^{\max} - x_2^{\min}}{2} = \frac{120-60}{2} = 30^\circ\text{C};$$

$$\Delta x_3 = \frac{x_3^{\max} - x_3^{\min}}{2} = \frac{150-120}{2} = 15 \text{ ат.}$$

Результаты кодирования представим в виде таблицы

Компоненты плана полного факторного эксперимента	Значения факторов					
	Концентрация		Температура		Давление	
	Физи- ческие	Кодиро- ванные	Физи- ческие	Кодиро- ванные	Физи- ческие	Кодиро- ванные
Центр плана	1,25	0	90	0	135	0
Интервал варьирования	0,25	1	30	1	15	1
Верхний уровень	1,5	+1	120	+1	150	+1
Нижний уровень	1,0	-1	60	-1	120	-1

4.2. Составление матрицы плана ПФЭ

Матрица трёхфакторного план строится аналогично (2.9):

$$Z = \begin{pmatrix} z_0 & z_1 & z_2 & z_3 \\ +1 & +1 & +1 & +1 \\ +1 & -1 & +1 & +1 \\ +1 & +1 & -1 & +1 \\ +1 & -1 & -1 & +1 \\ +1 & +1 & +1 & -1 \\ +1 & -1 & +1 & -1 \\ +1 & +1 & -1 & -1 \\ +1 & -1 & -1 & -1 \end{pmatrix} \Leftrightarrow Y = \begin{pmatrix} 0,580 \\ 0,508 \\ 0,500 \\ 0,503 \\ 0,590 \\ 0,503 \\ 0,480 \\ 0,510 \end{pmatrix}.$$

4.3. Вычисление коэффициентов по плану ПФЭ

Вычисление коэффициентов при факторах Z_1, Z_2, Z_3 производится по формуле (2.15):

$$b_0 = \frac{1}{8} \sum_{i=1}^8 Z_{1,0} y_i = \{(+) \cdot 0,580 + (+) \cdot 0,508 + (+) \cdot 0,500 +$$

$$+ (+) \cdot 0,503 + (+) \cdot 0,590 + (+) \cdot 0,530 + (+) \cdot 0,480 + (+) \cdot 0,510\} / 8 = 0,5251;$$

$$b_1 = \frac{1}{8} \sum_{i=1}^8 Z_{1,1} y_i = \{(+) \cdot 0,580 + (-) \cdot 0,508 + (+) \cdot 0,500 +$$

$$+ (-) \cdot 0,503 + (+) \cdot 0,590 + (-) \cdot 0,530 + (+) \cdot 0,480 + (-) \cdot 0,510\} / 8 = 0,01238;$$

$$b_2 = \frac{1}{8} \sum_{i=1}^8 Z_{1,2} y_i = \{(+) \cdot 0,580 + (+) \cdot 0,508 + (-) \cdot 0,500 +$$

$$+ (-) \cdot 0,503 + (+) \cdot 0,590 + (+) \cdot 0,530 + (-) \cdot 0,480 + (-) \cdot 0,510\} / 8 = 0,02688;$$

$$b_3 = \frac{1}{8} \sum_{i=1}^8 Z_{1,3} y_i = \{(+) \cdot 0,580 + (+) \cdot 0,508 + (+) \cdot 0,500 +$$

$$+ (+) \cdot 0,503 + (-) \cdot 0,590 + (-) \cdot 0,530 + (-) \cdot 0,480 + (-) \cdot 0,510\} / 8 = -0,002375.$$

Вычисление коэффициентов при парных произведениях производится по формуле (2.34):

$$b_{1,2} = \frac{1}{n} \sum_{i=1}^n Z_{1,1} Z_{1,2} y_i = \{(+) \cdot (+) \cdot 0,580 + (-) \cdot (+) \cdot 0,508 +$$

$$+ (+) \cdot (-) \cdot 0,500 + (-) \cdot (-) \cdot 0,503 + (+) \cdot (+) \cdot 0,590 + (-) \cdot (+) \cdot 0,530 +$$

$$+ (+) \cdot (-) \cdot 0,480 + (-) \cdot (-) \cdot 0,510\} / 8 = 0,02063;$$

$$b_{1,3} = \frac{1}{n} \sum_{i=1}^n Z_{1,1} Z_{1,3} y_i = \{(+) \cdot (+) \cdot 0,580 + (-) \cdot (+) \cdot 0,508 +$$

$$+ (+) \cdot (+) \cdot 0,500 + (-) \cdot (+) \cdot 0,503 + (+) \cdot (-) \cdot 0,590 + (-) \cdot (-) \cdot 0,530 +$$

$$+ (+) \cdot (-) \cdot 0,480 + (-) \cdot (-) \cdot 0,510\} / 8 = 0,004875;$$

$$b_{2,3} = \frac{1}{n} \sum_{i=1}^n Z_{1,2} Z_{1,3} y_i = \{(+) \cdot (+) \cdot 0,580 + (+) \cdot (+) \cdot 0,508 +$$

$$+ (-) \cdot (+) \cdot 0,500 + (-) \cdot (+) \cdot 0,503 + (+) \cdot (-) \cdot 0,590 + (+) \cdot (-) \cdot 0,530 +$$

$$+ (-) \cdot (-) \cdot 0,480 + (-) \cdot (-) \cdot 0,510\} / 8 = -0,005625.$$

В кодированном виде уравнение регрессии имеет вид:

$$y = 0,525 + 0,0124z_1 + 0,0269z_2 + 0,0206z_1z_2 + 0,00488z_1z_3 - 0,00563z_2z_3.$$

4.4. Проверка коэффициентов уравнения на значимость

Проверяется нулевая гипотеза H_0 об отсутствии различия между нулём и коэффициентом уравнения регрессии b_j . Сущность проверки гипотезы сводится к сравнению j -того коэффициента с его стандартным отклонением. Их отношение является опытным критерием Стьюдента, который сравнивается с табличным. В случае попадания опытного критерия Стьюдента в критическую область, нулевую гипотезу придётся отклонить и принять альтернативную $H_1: b_j \neq 0$. Проверку можно произвести и с помощью вероятного отклонения (см. разд. 2.4.).

Согласно (2.16) дисперсия коэффициентов

$$s_{b_j}^2 = \frac{1}{n} \cdot s_{on}^2 = \frac{1}{8} \cdot 0,65 \cdot 10^{-5} = 0,8125 \cdot 10^{-6}.$$

Квадратичное отклонение коэффициента b_j

$$s_{b_j} = \sqrt{0,8125 \cdot 10^{-6}} = \pm 0,0009014.$$

Опытный критерий Стьюдента по формуле (2.21)

$t_0 = 583$; $t_1 = 13,7$; $t_2 = 29,8$; $t_3 = 2,63$; $t_{1,2} = 22,9$; $t_{1,3} = 5,41$; $t_{2,3} = 6,24$.
Табличное значение двустороннего критерия Стьюдента для числа степеней свободы дисперсии воспроизводимости $\nu_{on} = 4$ и уровня значимости $\alpha = 0,05$ (см. табл. п.1) равно

$$t_4^{0,05} = 2,78.$$

Очевидно, что с доверительной вероятностью 95% нет оснований отвергнуть нулевую гипотезу $H_0: b_3 = 0$, т.е. коэффициент b_3 незначим. Гипотеза о равенстве нулю коэффициентов b_0 , b_1 , b_2 , $b_{1,2}$, $b_{1,3}$, $b_{2,3}$ противоречит результатам проверки, и с уровнем значимости $\alpha = 0,05$ нулевая гипотеза отклоняется и принимается альтернативная $H_1: b_j \neq 0$, т.е. коэффициенты b_0 , b_1 , b_2 , $b_{1,2}$, $b_{1,3}$, $b_{2,3}$ значимо отличаются от нуля.

Если для проверки коэффициентов на значимость воспользоваться доверительным отклонением (2.22), то

$$S_{b_j}^{0.05} = \pm 0,0009014 \cdot 2,78 = \pm 0,0025.$$

Сравнение вероятного отклонения с коэффициентами показывает, что ноль попадает в доверительные границы коэффициента b_3 , т.е. с вероятностью 95% коэффициент b_3 незначимо отличается от нуля. Это является основанием для исключения его из уравнения регрессии.

Таким образом, в кодированном виде искомое уравнение регрессии имеет вид:

$$y = 0,525 + 0,0124z_1 + 0,0269z_2 - 0,00238z_3 + \\ + 0,0206z_1z_2 + 0,00488z_1z_3 - 0,00563z_2z_3.$$

4.5. Проверка уравнения регрессии на адекватность

Расчётные значения функции отклика вычисляются по формуле

$$\hat{y}_i = b_0 + \sum_{j=1}^k b_j z_{i,j} + \sum_{\substack{l=1 \\ m=2 \\ l \neq m}}^k b_{l,m} z_{i,l} z_{i,m}.$$

Результаты расчётов представлены ниже:

$$Z = \begin{pmatrix} z_0 & z_1 & z_2 & z_3 \\ +1 & +1 & +1 & +1 \\ +1 & -1 & +1 & +1 \\ +1 & +1 & -1 & +1 \\ +1 & -1 & -1 & +1 \\ +1 & +1 & +1 & -1 \\ +1 & -1 & +1 & -1 \\ +1 & +1 & -1 & -1 \\ +1 & -1 & -1 & -1 \end{pmatrix} \Leftrightarrow \hat{Y} = \begin{pmatrix} 0,5843 \\ 0,5085 \\ 0,5005 \\ 0,5073 \\ 0,5858 \\ 0,5295 \\ 0,4795 \\ 0,5058 \end{pmatrix}.$$

Дисперсия адекватности $s^2_{ад}$ вычисляется по формуле (2.25):

$$s^2_{ад} = \frac{1}{v_{ад}} \cdot \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{8.6} = \frac{0,7325 \cdot 10^{-4}}{8.6} = 0,3663 \cdot 10^{-4}.$$

Опытный критерий Фишера вычисляется по формуле (2.27):

$$F_{\nu_{ад}, \nu_{оп}}^{оп} = \frac{S_{ад}^2}{S_{оп}^2},$$

где $\nu_{ад}=2$, $\nu_{оп}=4$;

$$F_{2.4}^{оп} = \frac{0,3663 \cdot 10^{-4}}{0,65 \cdot 10^{-5}} = 5,635.$$

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

$$F_{2.4}^{0.05} = 6,94 > F_{2.4}^{оп} = 5,635.$$

Опытный критерий Фишера меньше табличного значения, взятого для уровня значимости $\alpha=0,05$, т.е. нет оснований отклонить нулевую гипотезу об отсутствии различия между дисперсией воспроизводимости и дисперсией адекватности. Это значит, что уравнение регрессии адекватно.

4.6. Декодирование уравнения регрессии

Декодирование уравнения регрессии производится с помощью соотношения (2.4):

$$\begin{aligned} y &= b_0 + b_1 \left(\frac{x_1 - x_1^0}{\Delta x_1} \right) + b_2 \left(\frac{x_2 - x_2^0}{\Delta x_2} \right) + b_{1.2} \left(\frac{x_1 - x_1^0}{\Delta x_1} \right) \cdot \left(\frac{x_2 - x_2^0}{\Delta x_2} \right) + \\ &+ b_{1.3} \left(\frac{x_1 - x_1^0}{\Delta x_1} \right) \cdot \left(\frac{x_3 - x_3^0}{\Delta x_3} \right) + b_{2.3} \left(\frac{x_2 - x_2^0}{\Delta x_2} \right) \cdot \left(\frac{x_3 - x_3^0}{\Delta x_3} \right) = \\ &= \left(b_0 - b_1 \frac{x_1^0}{\Delta x_1} - b_2 \frac{x_2^0}{\Delta x_2} + b_{1.2} \frac{x_1^0}{\Delta x_1} \cdot \frac{x_2^0}{\Delta x_2} + b_{1.3} \frac{x_1^0}{\Delta x_1} \cdot \frac{x_3^0}{\Delta x_3} + \right. \\ &+ \left. b_{2.3} \frac{x_2^0}{\Delta x_2} \cdot \frac{x_3^0}{\Delta x_3} \right) + \left(\frac{b_1}{\Delta x_1} - \frac{b_{1.2} x_2^0}{\Delta x_1 \Delta x_2} - \frac{b_{1.3} x_3^0}{\Delta x_1 \Delta x_3} \right) \cdot x_1 + \\ &+ \left(\frac{b_2}{\Delta x_2} - \frac{b_{1.2} x_1^0}{\Delta x_1 \Delta x_2} - \frac{b_{2.3} x_3^0}{\Delta x_2 \Delta x_3} \right) \cdot x_2 + \left(- \frac{b_{1.3} x_1^0}{\Delta x_1 \Delta x_3} - \frac{b_{2.3} x_2^0}{\Delta x_2 \Delta x_3} \right) \cdot x_3 + \\ &+ \frac{b_{1.2}}{\Delta x_1 \Delta x_2} \cdot x_1 x_2 + \frac{b_{1.3}}{\Delta x_1 \Delta x_3} \cdot x_1 x_3 + \frac{b_{2.3}}{\Delta x_2 \Delta x_3} \cdot x_2 x_3. \end{aligned}$$

Подставляя в него значения коэффициентов b_j , $b_{1..m}$, координаты центра плана x^0_j и интервалы варьирования Δx_j , получим

$$y = 0,7595 - 0,3735x_1 - 0,0008542x_2 - 0,0005x_3 + \\ + 0,00275x_1x_2 + 0,0013x_1x_3 - 0,0000125x_2x_3.$$

После декодирования фактор x_3 появился в уравнении, и мы оказались в двойственном положении: с одной стороны, коэффициент b_3 незначим, но, с другой стороны, после декодирования коэффициент при давлении p появился в уравнении регрессии:

$$y = 0,7595 - 0,3735c_p - 0,0008542t - 0,0005p + \\ + 0,00275c_p t + 0,0013c_p p - 0,0000125tp.$$

Очевидно, что необходимо уточнение. Предположим, что фактор x_3 действительно не влияет на свойства цементного камня, и исключим коэффициенты $b_{1..3}$ и $b_{2..3}$. После пересчета коэффициентов мы получим следующее уравнение регрессии в кодированном виде:

$$y = 0,5251 + 0,01237z_1 + 0,02688z_2 + 0,02063z_1z_2.$$

Проверим его на адекватность. Вычисленная по формуле (2.25) дисперсия адекватности

$$s^2_{ад} = \frac{1}{\nu_{ад}} \cdot \sum_{i=1}^n (y_i - \hat{y}_i)^2 = \frac{0,5165 \cdot 10^{-3}}{8-4} = 0,1291 \cdot 10^{-3}.$$

Опытный критерий Фишера

$$F_{4,4}^{оп} = \frac{0,1291 \cdot 10^{-3}}{0,65 \cdot 10^{-5}} = 19,9,$$

где $\nu_{ад}=4$, $\nu_{оп}=4$ - числа степеней свободы дисперсии адекватности и дисперсии воспроизводимости соответственно. Опытный критерий Фишера сравниваем с табличным (табл. П.2):

$$F_{4,4}^{0,05} = 6,39 < F_{4,4}^{оп} = 19,9.$$

Опытный критерий Фишера больше табличного значения, взятого для уровня значимости $\alpha=0,05$, т.е. он попадает в критическую область, и

нет оснований принять нулевую гипотезу об отсутствии различия между дисперсией воспроизводимости и дисперсией адекватности. Это значит, что уравнение регрессии стало неадекватным эксперименту. Поскольку это явилось результатом окончательного удаления фактора x_3 из уравнения регрессии, значит, он влияет на процесс гидратации, и у нас нет оснований им пренебрегать.

Попробуем оставить коэффициент b_3 в уравнении регрессии:

$$y = 0,5252 + 0,01238z_1 + 0,02688z_2 - 0,002375z_3 + \\ + 0,02063z_1z_2 + 0,004875z_1z_3 - 0,005625z_2z_3.$$

Проверим его на адекватность. Вычисленная по формуле (2.25) дисперсия адекватности:

$$s^2_{ад} = \frac{1}{\nu_{ад}} \cdot \sum_{i=1}^n (y_i - \hat{y}_i)^2 = \frac{0,2813 \cdot 10^{-4}}{8-7} = 0,2813 \cdot 10^{-4}.$$

Опытный критерий Фишера

$$F_{1,4}^{оп} = \frac{0,2813 \cdot 10^{-4}}{0,65 \cdot 10^{-5}} = 4,33,$$

где $\nu_{ад}=1$, $\nu_{оп}=4$ - числа степеней свободы дисперсии адекватности и дисперсии воспроизводимости соответственно.

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

$$F_{1,4}^{0,05} = 7,71 > F_{1,4}^{оп} = 4,33.$$

Опытный критерий Фишера меньше табличного значения, т.е. с доверительной вероятностью $P=0,95$ уравнение регрессии адекватно.

После декодирования получим:

$$y = 0,7809 - 0,3735x_1 - 0,0008542x_2 - 0,0006583x_3 + \\ + 0,00275x_1x_2 + 0,0013x_1x_3 - 0,0000125x_2x_3.$$

В заключение необходимо отметить, что, вернув незначимый коэффициент b_3 в уравнение, мы нарушили правила регрессионного анализа, и неопределённость с влиянием фактора x_3 осталась. Для устранения этой неопределённости необходимо повторить эксперимент с большей точностью.

БИБЛИОГРАФИЧЕСКИЙ СПИСОК

1. Математический энциклопедический словарь. М.: Сов. энциклопедия, 1988. - 847 с.
2. Атназарова 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. Изнатов В.И. Организация и проведение эксперимента в бурении. М.: Недра, 1978, 94 с.
6. Кендалл м.Дж., Стьюарт А. Статистические выводы и связи: Пер. с англ. - М.: Наука, - 1973, - 900 с.
7. Кендалл М.Дж., Стьюарт А. Теория распределений: Пер с англ. - М.: Наука, 1966. - 588 с.
8. Крамер Г. Математические методы статистики: Пер. с англ. Изд. 2-е. - М.: Мир, 1975. - 648 с.
9. Линник Ю.В. Метод наименьших квадратов и основы математико-статистической теории обработки наблюдений. Изд. 2-е. - М., 1962.
10. Львовский Е.Н. Статистические методы построения эмпирических формул: Учеб. пособ. - М.: Высш. шк., 1982. - 224 с.
11. Налимов В.В., Чернова Н.А. Статистические методы планирования экстремальных экспериментов. М.: Наука, 1965. 340 с.
12. Саутин С.Н. Планирование эксперимента в химии и химической технологии. Л.: Химия, 1975. 48 с.
13. Хьютсон А. Дисперсионный анализ. - М.: Статистика, 1971.
14. Шеффе Д. Дисперсионный анализ. - М.: Физматгиз, 1963.
15. Finney D.J. The frakctional replication of factorial experiments// Annals of Eugenics. 1945. 12. №4. 291.
16. Box G.E.P., Hunter J.S. Multifactor experimental disigns for exploring response surfaces// Annals of mathematical statistics. 1957. 28. №1. 195.
17. Цивинский Д.Н. Применение метода пассивного эксперимента в нефтегазовом деле.: Учеб. пособ. Самар. гос. техн. ун-т. Самара, 2001, 84 с.

18. Зайдель А. Н. Погрешности измерений физических величин. Л.: Наука, 1985.
19. Лаудон Т. ЭВМ и машинные методы в геологии/ Пер. с англ. Д. А. Родионова. - М.: Мир, 1981. - 318 с.
20. 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.
21. Справочник по прикладной статистике/ Под ред. Э. Ллойда, У. Ледермана. Пер. с англ. С. А. Айвазяна и Ю. Н. Тюрина. М.: Финансы и статистика, 1990, В 2-х т.

**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	3,920
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	1,337	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

Критические значения распределения Фишера F_{α} для уровня значимости $\alpha=0,05$

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

Критические значения распределения Фишера F_{α} для уровня значимости $\alpha=0,025$

ν_2	Число степеней свободы числителя, ν_1															60	80	120	∞
	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30				
1	4052	499,5	5403	5625	5764	5859	5928	5981	6022	6056	6106	6157	6209	6265	6321	6387	6413	6396	
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	
3	34,12	30,82	29,46	28,71	28,24	27,91	27,67	27,49	27,35	27,25	27,17	27,10	27,04	27,00	26,97	26,94	26,92	26,13	
4	21,20	18,00	16,69	15,98	15,52	15,21	14,98	14,80	14,66	14,55	14,47	14,40	14,34	14,32	14,31	14,31	14,31	13,46	
5	16,26	13,27	12,06	11,39	10,97	10,67	10,46	10,29	10,16	10,05	9,99	9,92	9,85	9,77	9,71	9,66	9,65	9,11	
6	13,75	10,92	9,79	9,15	8,75	8,47	8,26	8,10	7,98	7,87	7,77	7,69	7,60	7,51	7,44	7,39	7,38	6,86	
7	12,25	9,55	8,45	7,85	7,46	7,17	6,99	6,83	6,72	6,62	6,54	6,47	6,39	6,30	6,23	6,18	6,17	5,65	
8	11,26	8,65	7,59	7,01	6,63	6,37	6,19	6,03	5,91	5,81	5,73	5,66	5,58	5,50	5,43	5,38	5,37	4,86	
9	10,56	8,02	6,99	6,42	6,06	5,80	5,61	5,47	5,36	5,28	5,11	4,96	4,81	4,73	4,66	4,57	4,48	4,31	
10	10,04	7,56	6,55	5,99	5,64	5,39	5,20	5,05	4,94	4,85	4,71	4,56	4,41	4,32	4,25	4,17	4,08	4,00	
11	9,65	7,21	6,22	5,67	5,32	5,07	4,88	4,74	4,63	4,54	4,40	4,25	4,10	4,02	3,94	3,86	3,78	3,69	
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	
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	
14	8,86	6,51	5,55	5,04	4,69	4,45	4,28	4,14	4,03	3,94	3,80	3,66	3,51	3,43	3,35	3,27	3,19	3,09	
15	8,69	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	
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	
17	8,40	6,11	5,18	4,67	4,34	4,10	3,93	3,79	3,68	3,59	3,45	3,31	3,16	3,08	3,00	2,92	2,83	2,75	
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	
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	2,59	
20	8,10	5,85	4,94	4,43	4,10	3,87	3,70	3,56	3,45	3,37	3,23	3,09	2,94	2,86	2,78	2,69	2,61	2,52	
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	
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	2,59	2,45	2,40	
23	7,89	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	2,36	
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	
25	7,77	5,57	4,68	4,19	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	
26	7,72	5,53	4,64	4,14	3,82	3,59	3,42	3,28	3,18	3,09	2,95	2,81	2,66	2,58	2,50	2,41	2,32	2,23	
27	7,68	5,49	4,60	4,11	3,79	3,56	3,39	3,25	3,15	3,06	2,93	2,79	2,63	2,55	2,47	2,38	2,29	2,19	
28	7,64	5,45	4,57	4,07	3,75	3,53	3,36	3,22	3,12	3,03	2,90	2,75	2,60	2,52	2,44	2,35	2,26	2,16	
29	7,60	5,42	4,54	4,04	3,73	3,50	3,33	3,19	3,09	3,00	2,87	2,73	2,57	2,49	2,41	2,32	2,23	2,14	
30	7,56	5,39	4,51	4,02	3,70	3,47	3,30	3,17	3,07	2,99	2,84	2,70	2,55	2,47	2,39	2,30	2,21	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	1,80	
60	7,09	4,96	4,10	3,63	3,31	3,10	2,93	2,80	2,70	2,61	2,47	2,33	2,18	2,10	2,03	1,94	1,84	1,60	
100	6,85	4,73	3,87	3,40	3,09	2,88	2,71	2,60	2,50	2,41	2,27	2,13	1,98	1,90	1,81	1,71	1,61	1,36	
∞	6,65	4,61	3,75	3,28	2,97	2,76	2,59	2,48	2,39	2,30	2,16	2,04	1,89	1,79	1,70	1,59	1,47	1,32	

Критические значения распределения Филера F_{α} для уровня значимости $\alpha=0,001$

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
1	4053*	5004	5404*	5625*	5764*	5850*	5920*	5981*	6023*	6056*	6107*	6159*	6209*	6256*	6291*	6327*	6361*	6396*
2	999,5	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
3	167,0	146,	141,1	137,1	134,6	132,8	131,6	130,6	129,9	129,2	128,5	127,4	126,4	125,9	125,4	125,0	124,5	123,5
4	74,14	61,25	56,18	53,44	51,71	50,53	49,66	49,00	48,47	48,03	47,41	46,74	46,10	45,74	45,43	45,09	44,75	44,40
5	47,18	37,12	33,20	31,09	29,75	28,64	28,16	27,64	27,24	26,92	26,42	25,91	25,39	25,14	24,87	24,60	24,33	23,79
6	36,51	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	15,75
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	12,78	12,53	12,32	12,12	11,70
8	25,42	18,48	15,89	14,36	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	9,33
9	22,86	16,30	13,90	12,50	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	7,81
10	21,04	14,91	12,55	11,28	10,49	9,92	9,52	9,20	8,96	8,75	8,45	8,13	7,80	7,64	7,47	7,30	7,12	6,76
11	19,69	13,61	11,55	10,25	9,56	9,05	8,66	8,35	8,12	7,92	7,63	7,32	7,01	6,85	6,68	6,50	6,35	6,00
12	18,64	12,91	10,90	9,65	8,95	8,46	8,06	7,71	7,48	7,29	7,00	6,71	6,40	6,23	6,09	5,93	5,78	5,42
13	17,81	12,31	10,21	9,07	8,35	7,86	7,43	7,09	6,86	6,69	6,40	6,13	5,85	5,78	5,63	5,47	5,30	4,97
14	17,14	11,78	9,73	8,62	7,92	7,43	7,00	6,66	6,43	6,26	6,00	5,72	5,45	5,41	5,25	5,10	4,94	4,60
15	16,59	11,34	9,34	8,25	7,57	7,09	6,74	6,47	6,26	6,09	5,81	5,54	5,25	5,10	4,95	4,80	4,64	4,31
16	16,12	10,97	9,00	7,94	7,27	6,81	6,46	6,19	5,99	5,81	5,55	5,27	4,99	4,85	4,70	4,54	4,39	4,06
17	15,76	10,66	8,73	7,69	7,02	6,56	6,22	5,95	5,76	5,59	5,32	5,05	4,78	4,63	4,48	4,33	4,18	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	3,67
19	15,08	10,16	8,28	7,25	6,62	6,16	5,85	5,59	5,39	5,22	4,97	4,70	4,43	4,29	4,14	3,99	3,84	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	3,38
21	14,59	9,77	7,90	6,91	6,28	5,85	5,56	5,31	5,11	4,95	4,70	4,44	4,17	4,03	3,88	3,74	3,58	3,26
22	14,39	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	3,15
23	14,19	9,47	7,69	6,69	6,08	5,65	5,33	5,08	4,89	4,73	4,48	4,23	3,96	3,82	3,68	3,53	3,38	3,05
24	14,03	9,34	7,55	6,55	5,94	5,50	5,23	4,99	4,80	4,64	4,39	4,14	3,87	3,74	3,59	3,45	3,29	2,97
25	13,88	9,22	7,45	6,45	5,84	5,40	5,15	4,91	4,71	4,56	4,31	4,06	3,79	3,65	3,52	3,37	3,22	2,89
26	13,74	9,12	7,36	6,41	5,80	5,36	5,07	4,83	4,64	4,49	4,24	3,99	3,72	3,59	3,44	3,30	3,15	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	2,75
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	2,69
29	13,39	8,86	7,12	6,19	5,61	5,19	4,87	4,64	4,45	4,29	4,05	3,80	3,54	3,41	3,27	3,12	2,97	2,64
30	13,29	8,77	7,05	6,12	5,53	5,12	4,82	4,59	4,40	4,24	4,00	3,75	3,49	3,36	3,22	3,07	2,92	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	2,23
60	11,36	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	1,89
120	11,36	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	1,89
∞	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	1,68	1,45

* Эти значения надо умножить на 100

ОГЛАВЛЕНИЕ

Условные обозначения	3
ВВЕДЕНИЕ	5
1. ПОСТАНОВКА ЗАДАЧИ	7
1.1. Обоснование метода наименьших квадратов	42
2. ПОЛНЫЙ ФАКТОРНЫЙ ЭКСПЕРИМЕНТ	49
2.1. Кодирование переменных	49
2.2. Построение плана экспериментов	51
2.3. Вычисление коэффициентов по плану ПФЭ	55
2.4. Проверка коэффициентов уравнения на значимость	56
2.5. Проверка уравнения регрессии на адекватность	59
2.6. Декодирование уравнения регрессии	62
2.7. Расширение полного факторного эксперимента	63
3. ДРОБНЫЙ ФАКТОРНЫЙ ЭКСПЕРИМЕНТ	66
4. ПРИМЕР РАСЧЁТА	72
4.1. Кодирование переменных	72
4.2. Составление матрицы плана ПФЭ	73
4.3. Вычисление коэффициентов по плану ПФЭ	74
4.4. Проверка коэффициентов уравнения на значимость	75
4.5. Проверка уравнения регрессии на адекватность	76
4.6. Декодирование уравнения регрессии	77
<i>БИБЛИОГРАФИЧЕСКИЙ СПИСОК</i>	80
<i>ПРИЛОЖЕНИЯ</i>	82

**Применение метода
полного факторного эксперимента
в нефтегазовом деле**

Цивинский Дмитрий Николаевич

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

Л. Р. №020595 от 09.07.97.

Подп. в печать 24.05.01.

Формат 60*84 1/16. Бум. типогр. №2.

Печать офсетная,

Усл. кр.-отг. 4,88. Уч.-изд.л. 4,65.

Тираж 170 экз. С-341.

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