

НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
МОСКОВСКИЙ ГОСУДАРСТВЕННЫЙ

**СТРОИТЕЛЬНЫЙ
УНИВЕРСИТЕТ**

Издательство МИСИ – МГСУ

СТАТИСТИЧЕСКИЕ МЕТОДЫ РЕШЕНИЯ ТЕХНОЛОГИЧЕСКИХ ЗАДАЧ



Министерство образования и науки Российской Федерации

НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
МОСКОВСКИЙ ГОСУДАРСТВЕННЫЙ СТРОИТЕЛЬНЫЙ УНИВЕРСИТЕТ

СТАТИСТИЧЕСКИЕ МЕТОДЫ РЕШЕНИЯ ТЕХНОЛОГИЧЕСКИХ ЗАДАЧ

*Рекомендовано Учебно-методическим объединением вузов РФ
по образованию в области строительства
в качестве учебного пособия для подготовки магистров
по направлению 08.04.01 Строительство
(магистерская программа «Строительное материаловедение»)
(24.02.2015 г., № 102-15/875)*

2-е издание (электронное)

Москва
2017

УДК 691.3:311
ББК 38.3
С78

Рецензенты:

доктор технических наук *Л.А. Алимов*, профессор кафедры технологии
вяжущих веществ и бетонов НИУ МГСУ;
доктор технических наук, профессор *А.Ф. Бурьянов*, исполнительный директор
Российской гипсовой ассоциации;
кандидат технических наук *И.В. Бессонов*, ведущий научный сотрудник
НИИСФ РААСН

Авторы-составители:

О.В. Александрова, Т.А. Мацевич, Л.В. Кирьянова, В.Г. Соловьев

С78 **Статистические методы решения технологических задач** [Электронный ресурс] : учебное пособие / [О.В. Александрова и др.] ; М-во образования и науки Рос. Федерации, Нац. исследоват. Моск. гос. строит. ун-т. — 2-е изд. (эл.). — Электрон. текстовые дан. (1 файл pdf : 154 с.). — Мсква. : Издательство МИСИ—МГСУ, 2017. — Систем. требования: Adobe Reader XI либо Adobe Digital Editions 4.5 ; экран 10".

ISBN978-5-7264-1645-8

Освещены вопросы планирования и обработки результатов эксперимента в области строительных материалов. Изложены процедуры математической обработки для оценки результатов эксперимента, проверки статистических гипотез. Рассмотрены вопросы планирования эксперимента с целью математического описания и выявления важнейших факторов, воздействующих на объект исследований с области строительных материалов. Приведены результаты исследований, связанные с определением свойств строительных материалов.

Для обучающихся по направлению 08.04.01 Строительство (магистерская программа «Строительное материаловедение»).

УДК 691.3:311
ББК 38.3

Деривативное электронное издание на основе печатного издания: Статистические методы решения технологических задач : учебное пособие / [О.В. Александрова и др.] ; М-во образования и науки Рос. Федерации, Моск. гос. строит. ун-т. — Москва. : Издательство МИСИ—МГСУ, 2015. — 160 с. — ISBN 978-5-7264-1076-0.

В соответствии со ст. 1299 и 1301 ГК РФ при устранении ограничений, установленных техническими средствами защиты авторских прав, правообладатель вправе требовать от нарушителя возмещения убытков или выплаты компенсации.

Глава 1. ОСНОВНЫЕ ПОНЯТИЯ ТЕОРИИ ВЕРОЯТНОСТЕЙ И МАТЕМАТИЧЕСКОЙ СТАТИСТИКИ

1.1. Случайная величина и ее закон распределения

Анализ значений величин, полученных в результате эксперимента, производится исходя из основных понятий, теорем и методов теории вероятностей и математической статистики.

Теория вероятностей — раздел математики, в котором изучаются закономерности случайных явлений.

Случайная величина — переменная величина, численные значения которой зависят от результата опыта. Обычно случайные величины обозначают большими латинскими буквами X, Y, Z и т.п., а возможные значения случайных величин — x_i, y_i, z_i .

На практике, как правило, используют случайные величины двух типов — дискретные и непрерывные.

Случайная величина называется дискретной, если множество ее значений конечно или счетно (т.е. множество бесконечное, но элементы его можно пронумеровать). Дискретность (от латин. *discretus* — разделенный, прерывистый) — прерывность.

Пример: дискретная случайная величина X — число отказавших элементов в приборе.

Законом распределения дискретной случайной величины называется правило, по которому каждому возможному значению случайной величины ставится в соответствие вероятность, с которой случайная величина может принять это значение. Закон распределения дискретной случайной величины может быть задан:

• *таблицей (рядом распределения)*. В первой строке таблицы находятся значения дискретной случайной величины, во второй — соответствующие вероятности. Данный ряд распределения может быть обозначен формулой

X	x_1	x_2	...	x_n
P	p_1	p_2	...	p_n

$$\sum_{i=1}^n P_i = 1;$$

• *графически (многоугольником распределения)*. По оси OX откладываются возможные значения дискретной случайной величины,

по оси OY — вероятности этих значений, и для наглядности полученные точки соединяются отрезками (рис. 1.1);

• *аналитически (функцией распределения)*. Функция распределения случайной величины X — функция $F(x)$, равная для любого значения x вероятности того, что случайная величина X примет значение меньше, чем x :

$$F(x) = P(X < x).$$

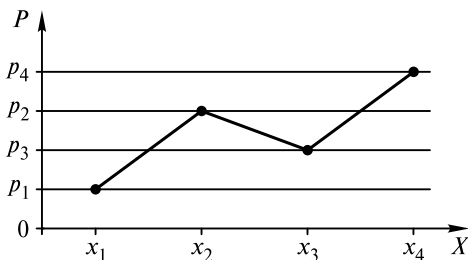


Рис. 1.1. Графическое обозначение дискретной случайной величины

$F(x)$ иногда называют интегральной функцией распределения, или интегральным законом распределения. Графиком $F(x)$ дискретной случайной величины является ступенчатая функция (рис. 1.2).

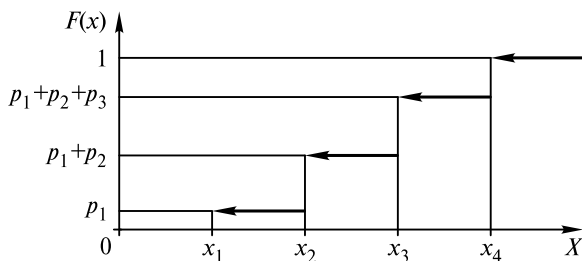


Рис. 1.2. Ступенчатая функция дискретной случайной величины

Случайная величина называется *непрерывной*, если множество ее значений целиком заполняет некоторый интервал. Например, непрерывная случайная величина Y — время безотказной работы прибора.

Непрерывная случайная величина может быть задана функцией распределения $F(x)$ или плотностью распределения $f(x)$.

Плотностью распределения непрерывной случайной величины X называется производная от функции распределения

$$f(x) = F'(x).$$

$f(x)$ иногда называется дифференциальной функцией распределения, или дифференциальным законом распределения. Кривая, являющаяся графиком плотности распределения, называется кривой распределения этой случайной величины (рис. 1.3).

При описании непрерывной случайной величины часто используют так называемые квантили. *Квантилем*, отвечающим заданной вероятности p , называют такое значение $x = x_p$, при котором функция распределения принимает значение, равное p (рис. 1.4), т.е.

$$F(x_p) = p.$$

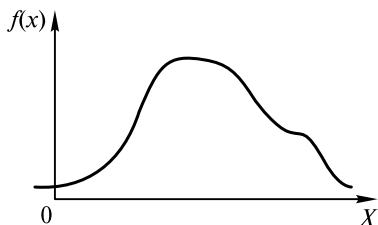


Рис. 1.3. Кривая распределения случайной величины

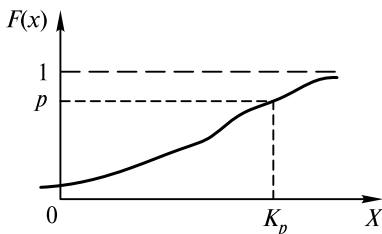


Рис. 1.4. Квантиль x с заданной вероятностью p

Некоторые квантили имеют особые названия. Например, квантиль, отвечающий значению $p = 0,5$, называют *медианой* распределения M_e . Медиана используется в качестве характеристики центра распределения. Квантили, соответствующие значениям $p = 0,25$ и $p = 0,75$, называют нижним и верхним квантилями (от латин. *quarta* — четверть). Зная значения достаточного числа квантилей, можно представить себе ход возрастания функции распределения.

Числовые характеристики *случайных величин* бывают двух видов:

- характеристики положения — математическое ожидание $M(X)$, мода M_0 ;
- характеристики рассеивания — дисперсия $D(X)$, среднеквадратическое отклонение $\sigma(X)$.

Математическое ожидание $M(X)$ для дискретной случайной величины X — это сумма произведений всех возможных значений дискретной случайной величины на вероятности этих значений:

$$M(X) = \sum_{i=1}^n x_i p_i.$$

Математическое ожидание $M(X)$ для непрерывной случайной величины X — это значение интеграла следующего вида:

$$M(X) = \int_{-\infty}^{+\infty} xf(x)dx.$$

Математическое ожидание можно воспринимать как некоторое значение, около которого группируются все возможные значения случайной величины.

Модой случайной величины M_0 называется наиболее вероятное значение этой случайной величины. Для дискретной случайной величины модой является то ее значение, у которого самая большая вероятность. Для непрерывной случайной величины модой является то ее значение, в котором плотность вероятности максимальная (рис. 1.5).

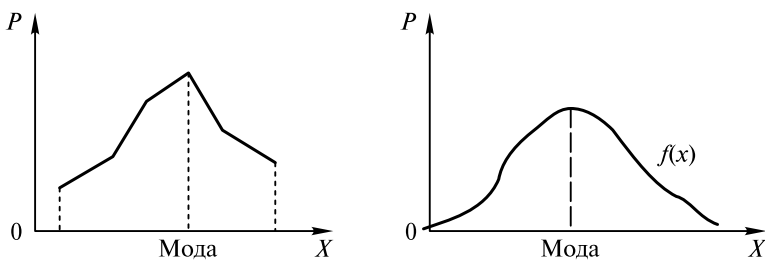


Рис. 1.5. Мода случайных величин

Дисперсией случайной величины $D(X)$ называют математическое ожидание квадрата отклонения этой случайной величины от ее математического ожидания:

$$D(X) = M\{[X - M(X)]^2\}.$$

Дисперсия (от латин. *dispersion* — рассеивание) — мера рассеивания, отклонения от среднего. На практике для вычисления дисперсии используют следующее ее свойство:

$$D(X) = M(X^2) - M^2(X),$$

т.е. для дискретной случайной величины

$$D(X) = \sum_{i=1}^n x_i^2 p_i - \left(\sum_{i=1}^n x_i p_i \right)^2,$$

а для непрерывной

$$D(X) = \int_{-\infty}^{+\infty} x^2 f(x) dx - \left(\int_{-\infty}^{+\infty} x f(x) dx \right)^2.$$

Среднеквадратическим отклонением случайной величины $\sigma(X)$ называют корень квадратный из дисперсии

$$\sigma(X) = \sqrt{D(X)}.$$

Коэффициент асимметрии определяется отношением

$$\gamma = \frac{M(X - M(X))^3}{(\sigma(X))^3}.$$

Если коэффициент асимметрии положителен, более «длинная» часть кривой плотности распределения лежит правее моды, что видно на рис. 1.6.

Если коэффициент асимметрии отрицателен, более «длинная» часть кривой плотности распределения лежит левее моды, что также видно на рис. 1.6.

Остановимся подробнее на наиболее часто встречающемся на практике законе распределения непрерывной случайной величины, который является предельным законом (т.е. к нему приближаются другие законы распределения при часто встречающихся типичных условиях), — нормальном законе распределения.

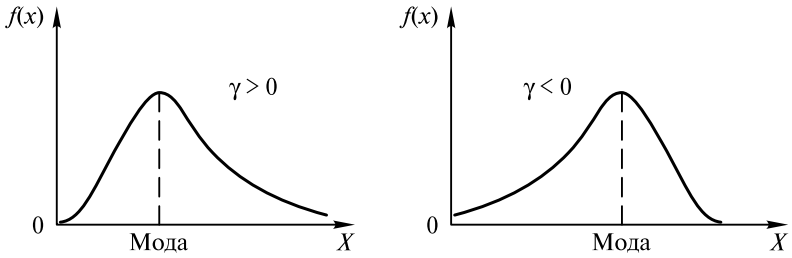


Рис. 1.6. Коэффициенты асимметрии для различных распределений

Непрерывная случайная величина называется распределенной по *нормальному закону* (или *закону Гаусса*), если ее плотность вероятности имеет следующий вид:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-m)^2}{2\sigma^2}}.$$

Кривая плотности распределения нормально распределенной случайной величины также называется *кривой Гаусса* (рис. 1.7).

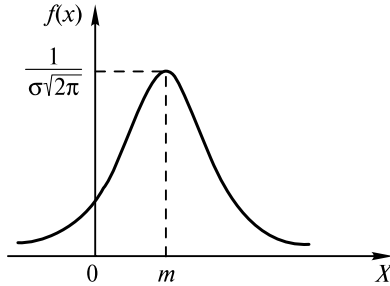


Рис. 1.7. Кривая Гаусса

Прямая $x = m$ является осью симметрии для кривой Гаусса. Параметр m не влияет на форму кривой, он определяет сдвиг по оси OX . Параметр σ определяет растяжение (или сжатие) кривой: чем больше значение σ , тем более пологая кривая.

Если $m = 0$, $\sigma = 1$, то кривая, как и соответствующее распределение, называется *нормированной*.

Для нормально распределенной случайной величины X математическое ожидание $M(X) = m$, а дисперсия $D(X) = \sigma^2$. Мода и меди-

ана совпадают с математическим ожиданием, коэффициент асимметрии равен нулю.

Вероятность попадания нормально распределенной случайной величины X в интервал от α до β

$$P(\alpha < X < \beta) = \Phi\left(\frac{\beta - m}{\sigma}\right) - \Phi\left(\frac{\alpha - m}{\sigma}\right),$$

где

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{-\frac{z^2}{2}} dz$$

— функция Лапласа. Значения функции Лапласа для положительных значений аргумента x табулированы, а для отрицательных x значение $\Phi(x)$ находят их условия ее нечетности

$$\Phi(-x) = -\Phi(x).$$

При $x \geq 5$ полагают, что $\Phi(-x) \approx 0,5$.

На практике часто используют «правило трех сигм», которое позволяет указать интервал практически возможных значений нормально распределенной случайной величины: если случайная величина распределена нормально, то абсолютная величина ее отклонения от математического ожидания не превосходит утроенного среднеквадратического отклонения с вероятностью, близкой к единице:

$$P(|X - M(X)| < 3\sigma) = 0,997 \approx 1.$$

Более подробно с введенными понятиями можно ознакомиться в [1–3].

1.2. Простейшие приемы статистического описания

Математическая статистика — раздел математики, в котором занимаются разработкой методов получения, описания, обработки опытных данных с целью изучения закономерностей случайных массовых явлений.

Множество значений случайной величины, полученных в результате эксперимента, представляет собой статистическую сово-

купность. *Генеральной статистической совокупностью* называется совокупность, содержащая в себе все возможные значения случайной величины. *Выборочной статистической совокупностью (выборкой)* называется совокупность, содержащая в себе только некоторую часть элементов генеральной совокупности. Число опытов n , содержащихся в выборке, называется *объемом выборки*.

Выборка должна быть *репрезентативной*, т.е. объекты выборки должны правильно представлять признаки генеральной совокупности. В силу закона больших чисел можно утверждать, что выборка будет репрезентативной, если ее осуществлять случайно.

Пусть имеется набор экспериментальных данных (выборка) объемом n . *Вариационным рядом* называют упорядоченные по возрастанию числовых значений элементы выборки $x_{(1)} \leq x_{(2)} \leq \dots \leq x_{(n)}$. Величина $x_{(k)}$ называется *k -й порядковой статистикой*. Крайние члены $x_{(1)} = x_{\min}$ и $x_{(n)} = x_{\max}$ называются *экстремальными* (соответственно минимальным и максимальным) *значениями выборки*. Промежуток между экстремальными значениями называют *интервалом варьирования*. Разность $x_{(n)} - x_{(1)}$ называется *размахом выборки*.

Выборочной медианой (серединой) выборки является величина

$$\widetilde{Me} = \begin{cases} x_{(m+1)} & \text{при } n = 2m + 1 \text{ (} n \text{ — нечетная величина);} \\ \frac{x_{(m)} + x_{(m+1)}}{2} & \text{при } n = 2m \text{ (} n \text{ — четная величина).} \end{cases}$$

Таким образом, выборочная медиана — это либо средний член вариационного ряда (если в выборке нечетное число данных), либо среднее арифметическое двух средних членов вариационного ряда (если в выборке четное число данных).

При статистическом анализе дискретной случайной величины используется простая таблица частот. Пусть выборка содержит k ($k \leq n$) различных значений и значение x_i встречается n_i раз, тогда величину n_i называют *частотой*, а значение x_i — *вариантой*. Сумма всех частот равна объему выборки

$$\sum_{i=1}^k n_i = n.$$

Множество пар (x_i, n_i) , где для каждой варианты указана ее частота, называют *статистическим рядом*, который записывают в виде простой таблицы частот (табл. 1.1).

Таблица частот

№	Значение	Частота
1	x_{\min}	n_1
...
m	x_{\max}	n_m

Если генеральная случайная величина — непрерывная, то таблица частот будет интервальной. Интервал варьирования разбивают на несколько, проводя группировку выборочных данных.

Число интервалов группировки (табл. 1.2) находится по эмпирической формуле (в которой округление идет до ближайшего целого):

$$k \approx 1 + 3,2 \cdot \lg n,$$

где n — объем выборки.

Длина интервала

$$h = \frac{x_{\max} - x_{\min}}{k},$$

где $x_{\max} - x_{\min}$ — размах выборки;

x_{\max} — максимальное значение выборки;

x_{\min} — минимальное значение выборки.

Таблица 1.2

Интервалы таблицы частот

№ интервала	Интервал	Частота
1	$(x_{\min}; x_{\max} + h)$	n_1
2	$(x_{\min} + h; x_{\max} + 2h)$	n_2
...
k	$(x_{\min} - h; x_{\max})$	n_k

Примечание. Частота n_k — количество данных, попавших в k -й интервал.

Полигоном частот называется ломаная, концы отрезков которой имеют координаты $\left(x_i; \frac{n_i}{n}\right)$. Для непрерывных величин полигон

представляет собой ломаную, которая проходит через точки, абсциссы которых являются серединами интервалов группировки (рис. 1.8).

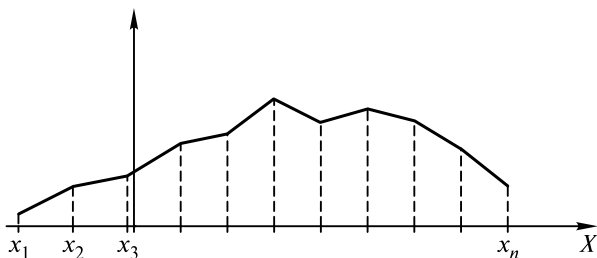


Рис. 1.8. Полигон частот

Эмпирическая функция распределения (рис. 1.9)

$$F_n(x) = \frac{v_n(x)}{n},$$

где $v_n(x)$ — число элементов выборки, оказавшихся меньше x . Функция $F_n(x)$ при большом числе наблюдений n близка в каждой точке x к теоретической функции распределения $F(x)$ наблюдаемой случайной величины X [о функции $F_n(x)$ говорят как о статистическом аналоге для $F(x)$].

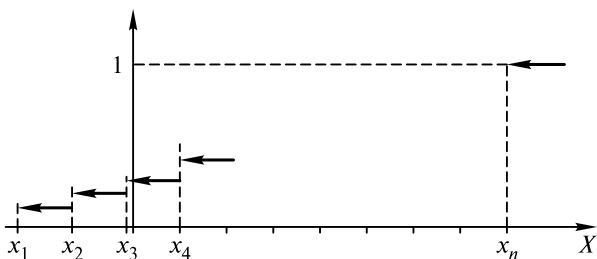


Рис. 1.9. Эмпирическая функция распределения

Гистограммой называется объединение прямоугольников, полученных следующим образом: по оси OX от x_{\min} откладываются интервалы длины d до x_{\max} на каждом интервале строится прямоугольник с высотой $h_i = \frac{p_i}{d}$, где p_i — относительная частота попадания

в i -й интервал группировки, $p_i = \frac{n_i}{n}$. При увеличении объема выборки и уменьшении длины интервала гистограмма будет приближаться к кривой плотности распределения (рис. 1.10).

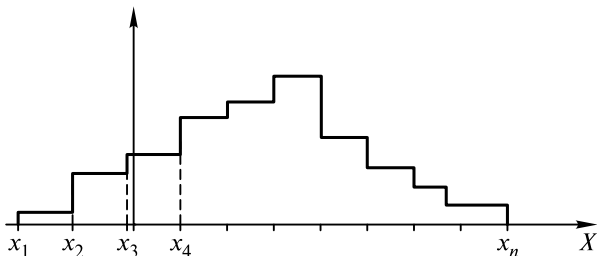


Рис. 1.10. Гистограмма функции распределения

Пример 1.1. В результате измерений активности цемента получено $n = 400$ значений, МПа. Значения активности цемента измеряются в диапазоне 28,5—55,5 МПа.

Требуется построить гистограмму активности цемента, соответствующую данному диапазону.

Решение

Рассчитываем значение числа интервалов по формуле

$$k = 1 + 3,2 \lg 4009.$$

Определяем длину интервала:

$$h = \frac{x_{\max} - x_{\min}}{k} = \frac{55,5 - 28,5}{9} = 30.$$

Далее определяем границы интервалов и заносим их во второй столбец табл. 1.3.

Таблица 1.3

Таблица распределения

№ интервала	Граница интервала	Середина интервала	Число наблюдений в интервале n_i
1	2	3	4
1	28,5—31,5	30,0	10
2	31,5—34,5	33,0	22
3	34,5—37,5	36,0	42

№ интервала	Граница интервала	Середина интервала	Число наблюдений в интервале n_i
1	2	3	4
4	37,5—40,5	39,0	56
5	40,5—43,5	42,0	86
6	43,5—46,5	45,0	77
7	46,5—49,5	48,0	65
8	49,5—52,5	51,0	22
9	52,5—55,5	54,0	20
Σ			400

В четвертый столбец таблицы заносим значения числа наблюдений в интервале n_i . По составленной таблице строим гистограмму (эмпирическую плотность распределения).

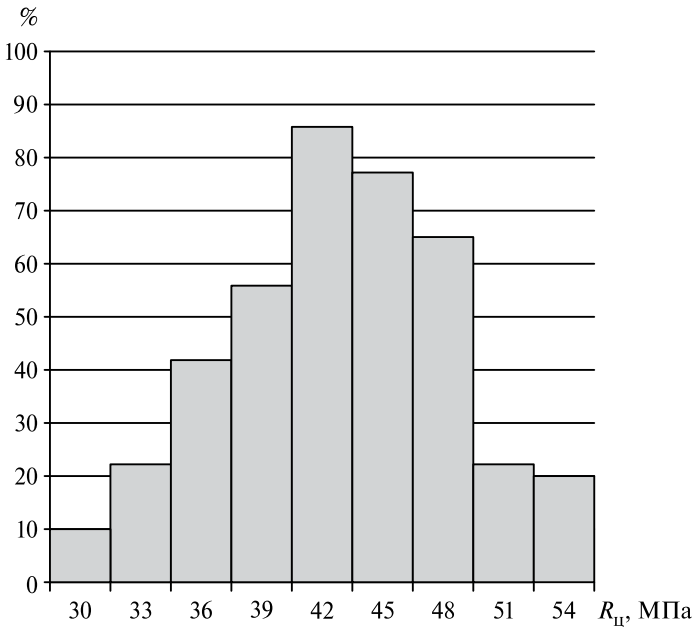


Рис. 1.11. Гистограмма активности цемента

На рис. 1.11 изображена гистограмма, соответствующая статистическому ряду, приведенному в табл. 1.3.

1.3. Точечные и интервальные оценки

1.3.1. Выборочные характеристики

Выборочные (эмпирические) характеристики, полученные при изучении выборки, считают *точечными оценками* соответствующих истинных характеристик генеральной совокупности. Так, например, выборочное среднее считают оценкой математического ожидания, выборочную дисперсию — оценкой дисперсии, и т.п.

Выборочное среднее

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

где k — число интервалов группировки;

x_i — значение варианты для дискретного случая или середина интервала группировки для непрерывного;

n_i — соответствующая интервалу частота.

В частности, если количество значений выборки совпадает с объемом выборки, то формула для выборочного среднего упрощается:

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

Выборочная мода — выборочное значение, которому соответствует наибольшая частота.

Выборочная дисперсия

$$s^2 = \frac{1}{n-1} \sum_{i=1}^k (x_i - \bar{x})^2 n_i = \frac{1}{n-1} \left(\sum_{i=1}^k x_i^2 n_i - n\bar{x}^2 \right), \quad (1.1)$$

где k — число интервалов группировки;

\bar{x} — выборочное среднее;

x_i — значение варианты для дискретного случая или середина интервала группировки для непрерывного;

n_i — соответствующая интервалу частота.

В частности, если количество различных значений выборки совпадает с объемом выборки, то формула для выборочной дисперсии упрощается:

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

Знаменатель выборочной дисперсии $n - 1$ называется числом степеней свободы. На практике чаще используется *выборочное среднеквадратическое отклонение*

$$s = \sqrt{s^2}.$$

С позиций практики, к оценке параметров предъявляются следующие требования: состоятельность, несмещенность и эффективность.

Оценка называется *состоятельной*, если при неограниченном увеличении объема выборки ее значение с вероятностью, равной единице, стремится к своему теоретическому значению.

Оценка называется *несмещенной*, если для любого объема выборки математическое ожидание равно оцениваемому параметру.

Несмещенная оценка называется *эффективной*, если среди всех оценок параметра она обладает наименьшей дисперсией.

Выборочное среднее есть состоятельная несмещенная оценка математического ожидания. Для выборки из нормальной генеральной совокупности эта оценка является также и эффективной.

Кроме точечных оценок неизвестных параметров распределения (математического ожидания, дисперсии, среднеквадратического отклонения и др.) используют и так называемые интервальные оценки.

Пусть найдено значение \bar{x} , которое является точечной оценкой неизвестного истинного значения x_0 . Задав достаточно большую вероятность p , определим такие два числа α и β , что

$$P\{\bar{x} - \alpha \leq x_0 \leq \bar{x} + \beta\} = p.$$

Тогда p называется *доверительной вероятностью*, или *уровнем надежности*, интервал $[\bar{x} - \alpha; \bar{x} + \beta]$ — *доверительным интервалом*, или *доверительной интервальной оценкой* истинного значения x_0 . Величина $q = 1 - p$ называется *уровнем значимости*. Уровень значимости, как и доверительная вероятность, задается до проведения расчетов.

Для симметричного относительно точечной оценки интервала *предельная ошибка выборки* Δ (или величина абсолютной погрешности оценивания) равна половине длины доверительного интервала, т.е. $\alpha = x_0 - \Delta$, $\beta = x_0 + \Delta$, тогда

$$\Delta = \frac{\beta - \alpha}{2}.$$

В общем случае предельная ошибка выборки равна наибольшему отклонению выборочного значения параметра от его истинного значения.

Для получения доверительного интервала наименьшей длины при заданном объеме выборки n и заданной доверительной вероятности p в качестве оценки параметра надо брать эффективную или асимптотически эффективную (т.е. становящуюся эффективной при больших n) оценку.

1.3.2. ДОВЕРИТЕЛЬНЫЙ ИНТЕРВАЛ ДЛЯ МАТЕМАТИЧЕСКОГО ОЖИДАНИЯ

Пусть изучается некоторая случайная величина X и требуется построить доверительный интервал для ее математического ожидания. Если объем выборки невелик, т.е. выборочных данных менее 30—40, то требуется дополнительное предположение о нормальном распределении этой случайной величины. Если же данных достаточно, то распределение изучаемой случайной величины может быть любым.

Обозначим:

n — объем выборки;

\bar{x} — выборочное среднее;

s — выборочное среднеквадратическое отклонение.

Доверительный интервал для математического ожидания $M(X)$ имеет вид

$$\bar{x} - \Delta \leq M(X) \leq \bar{x} + \Delta, \quad (1.2)$$

где Δ — предельная ошибка выборки, $\Delta = t_p \frac{s}{\sqrt{n}}$,

здесь t_p — критическое значение распределения Стьюдента, которое определяется по доверительной вероятности p из таблицы распре-

деления Стьюдента с $f = n - 1$ степенями свободы (табл. П1 приложения).

Заметим, что в некоторых учебниках приводятся таблицы, в которых t_p — квантиль уровня $(1 + p)/2$ распределения Стьюдента с $f = n - 1$ степенью свободы.

Пример 1.2. В результате эксперимента была определена прочность R_{28} твердения цемента, было выполнено шесть измерений прочности и получены значения, представленные в табл. 1.4.

Требуется рассчитать точечную оценку и доверительный интервал для математического ожидания прочности твердения цемента, если известно, что эта прочность имеет нормальное распределение.

Таблица 1.4

К примеру 1.2

Число опытов n	Активность R_{28} , МПа	Число опытов n	Активность R_{28} , МПа
1	45,0	4	46,0
2	45,2	5	46,6
3	45,8	6	48,8

Решение

В данном случае объем выборки $n = 6$.

Вычислим среднее арифметическое \bar{y} и оценку дисперсии s^2 по выборке:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i = \frac{1}{6} (45,0 + 45,2 + 45,8 + 46,0 + 46,6 + 48,8) = 46,23;$$

$$s^2 = \frac{1}{n-1} \left(\sum_{i=1}^n x_i^2 - n\bar{x}^2 \right) =$$

$$= \frac{1}{5} (45,0^2 + 45,2^2 + 45,8^2 + 46,0^2 + 46,6^2 + 48,8^2 - 6(46,23)^2) = 2,28.$$

Отсюда $s = \sqrt{s^2} = 1,5$ (МПа).

Зададимся доверительной вероятностью $p = 0,95$. Из табл. П1 приложения по величинам $p = 0,95$ и $f = n - 1 = 5$ найдем значение $t = 2,57$.

Подставляя найденные значения для \bar{x} , s , n и t в формулу (1.2), получим доверительный интервал для математического ожидания прочности R_{28} твердения цемента:

$$46,23 - \frac{2,57 \cdot 1,5}{6} \leq M(X) \leq 46,23 + \frac{2,57 \cdot 1,5}{6},$$

т.е. $45,59 \leq M(X) \leq 46,87$ (МПа).

1.3.3. ДОВЕРИТЕЛЬНЫЙ ИНТЕРВАЛ ДЛЯ СРЕДНЕКВАДРАТИЧЕСКОГО ОТКЛОНЕНИЯ

Доверительный интервал для среднеквадратического отклонения случайной величины может быть как односторонним, так и двусторонним. Односторонний интервал применяют, если дисперсия достаточно мала, т.е. близка к нулю. Так как среднеквадратическое отклонение не может быть отрицательным, то второй границей одностороннего интервала в данной ситуации автоматически становится число 0. В остальных же случаях целесообразнее применять двусторонний доверительный интервал.

Разница между односторонним и двусторонним доверительными интервалами состоит в следующем:

- Односторонний доверительный интервал для среднеквадратического отклонения

$$\sigma(X) \leq \sqrt{\frac{n-1}{h_p}} s,$$

где h_p — критическое значение распределения χ^2 , которое определяется с помощью таблицы распределения χ^2 с $f = n - 1$ степенями свободы по заданной надежности (доверительной вероятности) p исходя из условия $P(\chi^2(n-1) \geq h_p) = p$.

Значения h_p можно найти по табл. П3 приложения.

- Двусторонний доверительный интервал

$$\sqrt{\frac{n-1}{h_p''}} s \leq \sigma(X) \leq \sqrt{\frac{n-1}{h_p'}} s,$$

где h_p' и h_p'' определяются с помощью таблицы распределения χ^2 с $f = n - 1$ степенями свободы по заданной надежности p исходя из условий

$$P(\chi_i^2(n-1) > h'_p) = \frac{1+p}{2}, \quad P(\chi_{n-1}^2 > h''_p) = \frac{1-p}{2}.$$

Пример 1.3. На основе данных примера 1.2 требуется построить доверительный интервал для среднеквадратического отклонения прочности R_{28} твердения цемента с доверительной вероятностью $p = 0,9$.

Решение

Объем выборки $n = 6$. Уже подсчитано, что $s = 1,5$ (МПа).

По табл. П2 приложения для $f = n - 1 = 6 - 1 = 5$ находим $h_{0,9} = 1,61$. Тогда односторонний доверительный интервал для среднеквадратического отклонения прочности R_{28} твердения цемента будет следующим:

$$\sigma(X) \leq \sqrt{\frac{6-1}{1,61}} 1,5;$$

$$\sigma(X) \leq 2,643.$$

Теперь построим двусторонний доверительный интервал для среднеквадратического отклонения прочности R_{28} твердения цемента.

Сначала определим $\frac{1-p}{2} = 0,05$, $\frac{1+p}{2} = 0,95$. Далее по табл. П3 приложения для $f = 5$ находим $h'_{0,9} = h_{0,95} = 1,145$ и $h''_{0,9} = h_{0,05} = 11,07$. Тогда двусторонний доверительный интервал для среднеквадратического отклонения прочности R_{28} твердения цемента равен

$$\sqrt{\frac{6-1}{11,07}} 1,5 \leq \sigma(X) \leq \sqrt{\frac{6-1}{1,145}} 1,5,$$

или

$$1,01 \leq \sigma(X) \leq 3,14.$$

Не следует думать, что во всех случаях целесообразно задаваться возможно большей надежностью статистического вывода. Можно задать доверительную вероятность $p = 0,99$. Новое значение t , найденное из табл. П2 приложения, составит 4,03, а интервальная оценка примет вид $45,22 \leq M(X) \leq 47,24$ (МПа). Как и следовало ожидать, с большей надежностью можно гарантировать только бо-

лее широкий доверительный интервал для математического ожидания при тех же опытных данных.

Вопрос о выборе конкретного значения для доверительной вероятности решается исходя из требований разумности и точности. Ведь чем больше доверительная вероятность, тем шире доверительный интервал. Для того чтобы сузить доверительный интервал, т.е. сделать более точный статистический вывод об исследуемой случайной величине, требуется увеличить объем выборки. Часто проведение дополнительных измерений или опытов требует дополнительных затрат, и бывает необходимо уменьшить объем выборки, сохранив необходимую точность.

Для определения необходимого объема выборки при доверительном оценивании математического ожидания используется формула

$$n = \left(\frac{t_p}{\Delta} \right)^2 s^2. \quad (1.3)$$

Пример 1.4. На основе измерений активности цемента (см. пример 1.1) выявлено, что объем выборки $n = 400$.

Требуется найти необходимый объем выборки, при котором среднее отличалось бы от математического ожидания не более чем на $\Delta = 1,5$ МПа с доверительной вероятностью $p = 0,95$.

Решение

Для определения необходимого объема выборки (табл. 1.5) воспользуемся приведенной формулой (1.3). По данным примера 1.1 вычислим среднее \bar{x} и оценку дисперсии s^2 — формула (1.1).

Таблица 1.5

Матрица выборки

№ интервала	\bar{x}	x_i^*	m_i	$y_i^* \cdot m_i$	$(x_i^* - \bar{x})$	$(x_i^* - \bar{x})^2$	$m_i(x_i^* - \bar{x})^2$
1	42,8	30,0	10	300,0	-12,8	163,84	1638,40
2		33,0	22	726,0	-9,8	96,04	2112,88
3		36,0	42	1512,0	-6,8	46,24	1942,08
4		39,0	56	2184,0	-3,8	14,44	808,64
5		42,0	86	3612,0	-0,8	0,64	55,04
6		45,0	77	3465,0	2,2	4,84	372,68

№ интервала	\bar{x}	x_i^*	m_i	$y_i^* \cdot m_i$	$(x_i^* - \bar{x})$	$(x_i^* - \bar{x})^2$	$m_i(x_i^* - \bar{x})^2$
7	42,8	48,0	65	3120,0	5,2	27,04	1757,60
8		51,0	22	1122,0	8,2	67,24	1479,28
9		54,0	20	1080,0	11,2	125,44	2508,80
Σ							12675,40

$$\bar{x} = 42,8, s^2 = 31,7.$$

Тогда необходимый объем выборки

$$n = \left(\frac{t_\gamma}{\Delta} \right)^2 s^2 = \left(\frac{1,96}{1,5} \right)^2 31,7 \approx 54.$$

1.4. Основные положения теории проверки статистических гипотез

1.4.1. ГИПОТЕЗА В МАТЕМАТИЧЕСКОЙ СТАТИСТИКЕ

Процедура сопоставления рассматриваемых гипотез с выборкой и вынесения решения относительно приемлемости этих гипотез получила название *проверки гипотез*. Такая процедура используется всякий раз, когда необходим обоснованный вывод о преимуществах того или иного способа измерений, инвестиций, технологического процесса, об эффективности нового метода управления, производства чего-либо, о пользе удобрений, лекарств, об уровне доходности ценных бумаг, и т.д.

Гипотезой в математической статистике называется любое утверждение о виде или свойствах распределения вероятностей, лежащее в основе наблюдаемых явлений.

По своему содержанию статистические гипотезы можно подразделить на несколько основных типов:

- гипотезы о числовых значениях параметров случайной величины;
- гипотезы о виде распределения исследуемой случайной величины;
- гипотезы об общем виде модели, описывающей статистическую зависимость между случайными величинами.

Проверка статистической гипотезы заключается в том, чтобы оценить, можно ли считать случайным расхождение между выдвинутой гипотезой и результатами выборочного наблюдения. Такая оценка всегда носит вероятностный характер.

Если расхождение между эмпирическими и теоретическими значениями не выходит за пределы случайной ошибки, то можно считать, что с заданной вероятностью выдвинутая гипотеза не опровергается.

При этом справедливость самой гипотезы не доказывается, а лишь делается вывод о том, можно ли считать ее допустимой или необходимо опровергнуть.

Обычно выделяют некоторую основную (нулевую) гипотезу H_0 . Наряду с ней рассматривают конкурирующую (альтернативную) гипотезу H_1 , являющуюся логическим отрицанием основной.

Статистический критерий — это правило, по которому нужно принять или опровергнуть выдвинутую гипотезу. Таким образом, статистический критерий представляет собой такое правило, по которому устанавливается, при каких результатах выборочного исследования основная гипотеза не может быть отклонена, а при каких от нее необходимо отказаться.

Суть проверки статистической гипотезы заключается в следующем: используется специально составленная *статистика* (некоторая функция от выборки), точное или приближенное распределение которой известно. Множество возможных значений статистики делится на две взаимно дополняющих области:

- область отклонения выдвинутой гипотезы или критическая область критерия;
- область принятия гипотезы или область допустимых значений статистики.

При этом возможны четыре случая (рис. 1.12).

Вероятности ошибок 1-го и 2-го рода однозначно определяются выбором критической области.

Будем обозначать вероятность ошибки 1-го рода через q . Вероятность ошибки 2-го рода обозначим w .

Вероятность q отвергнуть гипотезу, когда она верна, называют *уровнем значимости критерия*, т.е. это вероятность совершить ошибку 1-го рода.

Обычно уровень значимости q принимают равным 0,05; 0,01; 0,005; 0,001.

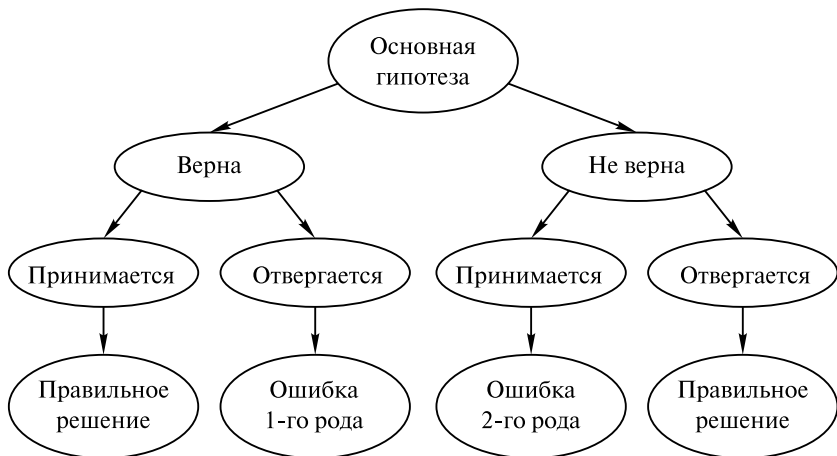


Рис. 1.12. К проверке статистической гипотезы

Вероятность $1 - w$ отвергнуть гипотезу, когда она не верна, т.е. не допустить ошибку 2-го рода, называют *мощностью критерия*.

Уменьшая уровень значимости q , мы снижаем вероятность появления ошибки 1-го рода. Однако, если основная гипотеза не верна, то, уменьшая q , мы увеличиваем область допустимых значений и, соответственно, вероятность появления ошибки второго рода w . Стремление увеличить мощность критерия $1 - w$ при неизменном объеме выборки приводит к расширению критической области, т.е. повышает вероятность ошибки первого рода q . Одновременно уменьшить вероятности q и w можно, лишь увеличивая объем выборки n . Если же объем выборки фиксирован, то при равных q выбирают тот критерий, где меньше w . То есть критическая область критерия должна быть такой, чтобы при заданном уровне значимости q мощность критерия $1 - w$ была максимальной.

Точка, разделяющая критическую область и область принятия нулевой гипотезы, называется критической.

При заданном уровне значимости критическая область (т.е. область отклонения выдвинутой гипотезы) может быть односторонняя (правосторонняя, левосторонняя) или двусторонняя в зависимости от вида альтернативной гипотезы (рис. 1.13).

Общий порядок проверки статистических гипотез следующий:

- Формулируются основная (проверяемая) и альтернативная гипотезы.



Рис. 1.13. Критические области

- Выбирается статистический критерий для проверки справедливости основной гипотезы.
- Определяются критическая область и область допустимых значений, значения статистики критерия в критических точках.
- По результатам статистического исследования подсчитывается фактическое значение статистики критерия.
- На основе сравнения фактического и критического значений статистики критерия делается вывод о правдоподобности или необходимости отклонения выдвинутой гипотезы.

Схема проверки статистической гипотезы не дает точного вывода о ее верности или неверности, так как принятие гипотезы всегда происходит на некотором субъективно принятом уровне надежности и основывается на значении статистики, построенной по конечной выборке.

Принятие основной гипотезы не означает, что она является единственно подходящей, просто она не противоречит выборочным данным. Однако таким же свойством могут обладать и другие гипотезы.

1.4.2. ГИПОТЕЗА О РАВЕНСТВЕ ДВУХ ДИСПЕРСИЙ ПРИ НЕИЗВЕСТНЫХ СРЕДНИХ (КРИТЕРИЙ ФИШЕРА — СНЕДЕКОРА)

Рассмотрим две генеральные случайные величины, имеющие нормальное распределение с неизвестными средними и дисперсиями. Требуется проверить гипотезу о равенстве дисперсий:

$$H_0: \sigma_1^2 = \sigma_2^2.$$

Приведем в виде табл. 1.6 критерии проверки гипотезы о равенстве неизвестных дисперсий.

Таблица 1.6

Критерии проверки гипотезы

Основная гипотеза H_0	Предположение	Статистика критерия	Альтернатива H_1	Область принятия основной гипотезы
$\sigma_1^2 = \sigma_2^2$	Математические ожидания m_1 и m_2 — неизвестны	$F_{\text{расч}} = \frac{s_1^2}{s_2^2}$, где $s_1^2 > s_2^2$	$\sigma_1^2 > \sigma_2^2$	$F_{\text{расч}} < F_q$
			$\sigma_1^2 \neq \sigma_2^2$	$F_{\text{расч}} < F_{q/2}$

Здесь F_q — критическое значение распределения Фишера — Снедекора с $f_1 = n_1 - 1$ и $f_2 = n_2 - 1$ степенями свободы.

Пример 1.5. При проверке размеров подшипников из двух поставленных разными заводами партий по 10 штук в каждой были обнаружены отклонения от номинала, характеризующиеся выборочными дисперсиями $s_1^2 = 9; s_2^2 = 8,5$.

Можно ли считать при уровне значимости $\alpha = 0,05$ одинаковой точность изготовления подшипников разными заводами?

Решение

Найдем значение статистики критерия $F_{\text{расч}} = 9/8,5 = 1,06$. Теперь по табл. П4 приложения находим критическое значение критерия Фишера — Снедекора $F_{0,05} = 3,18$. Значение статистики значительно меньше критического, следовательно, выборочные данные не противоречат выдвинутой гипотезе. Итак, точность изготовления подшипников разными заводами можно считать одинаковой при уровне значимости $q = 0,05$.

Пример 1.6. Для сравнения точности двух электровлагомеров ЭВ-2К каждым из них проведено 10 измерений влажности одного и того же участка доски. Результаты замеров влажности $W, \%$, первым и вторым приборами приведены в табл. 1.7.

Таблица 1.7

К примеру 1.6

Первый прибор	Второй прибор	Первый прибор	Второй прибор
35	42	43	40
41	39	42	37
47	42	37	44
38	43	49	44
41	45	39	43

Требуется вычислить значения средних и выборочных дисперсий для каждого прибора.

Решение

Вычисленные значения средних и выборочных дисперсий для каждого прибора соответственно равны $\bar{x}_1 = 41,2; \bar{x}_2 = 41,9; s_1^2 = 18,84; s_2^2 = 6,32$. Дисперсии отличаются почти втрое. Следует ли отсюда, что точность первого влагомера меньше, чем второго? Вычислим $F_{\text{расч}}$ по формуле

$$F_{\text{расч}} = s_1^2/s_2^2 = 1,84/6,32 = 2,98.$$

В данном случае в числителе должна быть дисперсия s_1^2 .

Зададимся уровнем значимости $q = 0,05$. Числа степеней свободы каждой из дисперсий равны $f_1 = f_2 = 10 - 1 = 9$. Из табл. П4 приложения для $q = 0,05; f_1 = f_2 = 9$ найдем $F_{0,05} = 3,18$.

Выявлено, что расчетное значение меньше табличного, но эта разница совсем незначительна и не дает оснований сделать вывод о значимости расхождения в точности исследуемых влагомеров по результатам данного эксперимента. Для окончательного решения вопроса необходимо повторить эксперимент, существенно увеличив объем каждой выборки.

1.4.3. ГИПОТЕЗА О РАВЕНСТВЕ СРЕДНИХ ДВУХ НОРМАЛЬНЫХ РАСПРЕДЕЛЕНИЙ

На практике часто встречается случай, когда средний результат одной серии экспериментов отличается от среднего результата другой серии. При этом возникает вопрос, можно ли объяснить данное расхождение случайными ошибками эксперимента или оно вызвано некоторыми закономерностями?

В финансовом анализе данная задача возникает при сопоставлении уровня доходности различных активов, в промышленности — при выборочном контроле качества изделий, изготовленных на разном оборудовании, или при исследовании надежности технических систем, где результаты измерений сравниваются с предыдущими результатами.

Рассмотрим две генеральные случайные величины, имеющие нормальное распределение, и проверим гипотезу о равенстве средних значений.

Приведем в виде табл. 1.8 все три критерия проверки гипотезы о равенстве средних при неизвестных дисперсиях.

В случае неодинаковых дисперсий также можно использовать t -критерий Стьюдента, причем число степеней свободы f определяют по формуле

$$f = \left\lfloor \frac{(s_1^2/n_1 + s_2^2/n_2)^2}{\frac{(s_1^2/n_1)^2}{n_1+1} + \frac{(s_2^2/n_2)^2}{n_2+1} - 2} \right\rfloor.$$

Здесь квадратные скобки означают, что найденное число требуется округлить до целого.

Таблица 1.8

Критерии проверки гипотезы о равенстве средних при неизвестных дисперсиях

Основная гипотеза H_0	Предположение	Статистика критерия	Альтернатива H_1	Область принятия основной гипотезы
$m_1 = m_2$	Дисперсии неизвестны, но равны, их оценки $s_1^2 = \frac{1}{n_1-1} \sum_{i=1}^n (X_i^{(1)} - \bar{X}^{(1)})^2$ $s_2^2 = \frac{1}{n_2-1} \sum_{i=1}^n (X_i^{(2)} - \bar{X}^{(2)})^2$	$t(n_1 + n_2 - 2) = \frac{\bar{x}^{(1)} - \bar{x}^{(2)}}{s \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$, где $s^2 = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1 + n_2 - 2}$	$m_1 > m_2$	$t < t_q$
			$m_1 < m_2$	$t > -t_q$
			$m_1 \neq m_2$	$ t < t_{q/2}$
$m_1 = m_2$	Дисперсии неизвестны, их оценки $s_1^2 = \frac{1}{n_1-1} \sum_{i=1}^n (X_i^{(1)} - \bar{X}^{(1)})^2$ $s_2^2 = \frac{1}{n_2-1} \sum_{i=1}^n (X_i^{(2)} - \bar{X}^{(2)})^2$	$t = \frac{ \bar{x}_1 - \bar{x}_2 }{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$	$m_1 \neq m_2$	$ t < t_{q/2}$

Примечание. t_q — критическое значение распределения Стьюдента с числом степеней свободы $f = n_1 + n_2 - 2$, которое определяется по табл. П2 приложения.

В случае принятия основной гипотезы говорят, что различие выборочных средних статистически не значимо и оценка общего математического ожидания есть

$$\tilde{m} = \frac{\bar{x}^{(1)}n_1 + \bar{x}^{(2)}n_2}{n_1 + n_2}.$$

Пример 1.7. На предприятии вводится новая технология производства. Расход сырья как по старой, так и по новой технологии имеет нормальное распределение (табл. 1.9).

Требуется выяснить, влияет ли технология на средний расход сырья. Принять $q = 0,05$.

Таблица 1.9

К примеру 1.10

Показатели	Технология						
	старая			новая			
Расход сырья	304	307	308	303	304	306	308
Число изделий	1	4	4	2	6	4	1

Решение

Найдем выборочные средние и выборочные дисперсии.

$$n_1 = 9; n_2 = 13; \bar{x}^{(1)} = 307,11; \bar{x}^{(2)} = 304,77; s_1^2 = 2,378; s_2^2 = 1,165.$$

Сначала проверим гипотезу о равенстве неизвестных дисперсий:

$$H_0: \sigma_1^2 = \sigma_2^2 \text{ против альтернативы } H_1: \sigma_1^2 > \sigma_2^2.$$

Используем статистику Фишера — Снедекора.

$$\text{Расчетное значение статистики } F_{\text{расч}} = \frac{2,378}{1,685} = 1,41.$$

По табл. П4 приложения находим критическое значение критерия Фишера — Снедекора $F_{0,05} = 2,85$.

Таким образом, значение статистики меньше критического значения и можно принять гипотезу о равенстве дисперсий.

Далее проверим гипотезу о равенстве неизвестных математических ожиданий: $H_0: m_1 = m_2$ против альтернативы $H_1: m_1 > m_2$.

Расчетное значение статистики критерия Стьюдента

$$t_{(20)} = \frac{307,11 - 304,77}{8 \cdot 2,378 + 12 \cdot 1,685} \cdot 20 = 1,9622$$

По табл. П1, П2 приложения находим значение границы критической области $t_{01} = 1,72$. Значение статистики критерия попало в критическую область, поэтому следует отвергнуть основную гипотезу и принять альтернативную. Можно сделать вывод, что технология влияет на расход сырья: при старой технологии в среднем сырья расходуется больше.

1.4.4. ГИПОТЕЗА О РАВЕНСТВЕ НЕСКОЛЬКИХ ДИСПЕРСИЙ ПРИ НЕИЗВЕСТНЫХ СРЕДНИХ (КРИТЕРИЙ БАРТЛЕТТА)

Проверка гипотезы о равенстве дисперсий величин

Пусть в результате наблюдения над k случайными величинами $X_{(1)}, X_{(2)}, \dots, X_{(k)}$ получено k групп данных:

$$x_{(1)1}, x_{(1)2}, \dots, x_{(1)n_1};$$

$$x_{(2)1}, x_{(2)2}, \dots, x_{(2)n_2};$$

...

$$x_{(k)1}, x_{(k)2}, \dots, x_{(k)n_k}.$$

Общее количество данных (объем выборки): $n = \sum_{i=1}^k n_i$.

Проверим гипотезу о равенстве дисперсий изучаемых величин $\sigma_{(1)}^2 = \sigma_{(2)}^2 = \dots = \sigma_{(k)}^2$ с помощью критерия Бартлетта.

С этой целью необходимо выполнить расчеты в следующей последовательности:

- Найти групповые средние:

$$\bar{x}_{(i)} = \frac{1}{n_i} \sum_{j=1}^{n_i} x_{(i)j}.$$

- Найти несмещенные оценки групповых дисперсий:

$$\tilde{s}_{(i)}^2 = \frac{1}{n_i - 1} \sum_{j=1}^{n_i} (x_{(i)j} - \bar{x}_{(i)})^2.$$

- Найти общую несмещенную оценку дисперсии:

$$\tilde{s}^2 = \frac{1}{n - k} \sum_{i=1}^k (n_i - 1) \tilde{s}_{(i)}^2.$$

- Вычислить статистику Бартлетта:

$$\mu = \frac{\sum_{i=1}^k (n_i - 1) \ln \frac{\tilde{s}^2}{\tilde{s}_{(i)}^2}}{1 + \frac{1}{3(k-1)} \left[\left(\sum_{i=1}^k \frac{1}{n_i - 1} \right) - \frac{1}{n - k} \right]}.$$

• По уровню значимости q найти критическое значение с помощью таблиц распределения χ^2 :

$$\mu_{кр} = \chi_{1-q, k-1}^2.$$

Если значение статистики μ меньше критического $\mu_{кр}$, то выборочные данные согласуются с гипотезой. Если значение статистики μ больше критического $\mu_{кр}$ (или равно ему), то выборочные данные противоречат высказанной гипотезе о равенстве групповых дисперсий.

Пример 1.8. В табл. 1.10 приведены данные по объемам работ, выполненных на стройке за смену для четырех бригад.

Требуется проверить по критерию Бартлетта гипотезу о равенстве групповых дисперсий. В этой задаче $k = 4$; $n_i = 4$, $i = 1; 2; 3; 4$. Объем выборки $n = 16$.

Решение

Расчеты по первым двум пунктам проведем в табл. 1.11.

Таблица 1.10

Таблица объема работ

Номер бригады i	Объем выполненной работы $x_{(ij)}$			
1	140	144	142	145
2	150	149	152	152
3	148	149	146	147
4	150	155	154	152

Таблица 1.11

Расчет дисперсии

Номер бригады i	Объем выполненной работы $x_{(ij)}$				Групповые средние $\bar{x}_{(i)}$	Групповые дисперсии $\tilde{s}_{(i)}^2$
1	140	144	142	145	142,75	4,92
2	150	149	152	150	150,25	1,58
3	148	149	146	147	147,50	1,67
4	150	155	154	152	152,75	4,92

Общая оценка дисперсии

$$\tilde{s}^2 = \frac{1}{16-4} \sum_{i=1}^4 3\tilde{s}_{(i)}^2 = 3,27.$$

Статистика Бартлетта:

$$\mu = \frac{\sum_{i=1}^4 3 \ln \frac{3,27}{\tilde{s}_{(i)}^2}}{1 + \frac{1}{9} \left[\left(\sum_{i=1}^4 \frac{1}{3} \right) - \frac{1}{16-4} \right]} = 1,538.$$

Выберем уровень значимости $q = 0,05$ и по таблице распределения χ^2 с 3 степенями свободы находим критическое значение: $\mu_{кр} = \chi_{0,95, 3}^2 = 7,82$.

Значение статистики меньше критического, следовательно, можно принять гипотезу о равенстве групповых дисперсий.

Проверка гипотезы о равенстве групповых средних

После того, как выполнена проверка условия равенства дисперсии (с помощью критерия Бартлетта), можно проводить проверку гипотезы о равенстве групповых средних.

Пусть основная гипотеза состоит в том, что все групповые средние равны между собой, т.е.

$$H_0: MX_{(1)} = MX_{(2)} = \dots = MX_{(k)}.$$

Проверку этой гипотезы проводят с помощью статистики

$$F = \frac{s_M^2 / (r-1)}{s_r^2 / (n-r)}, \quad (1.4)$$

которую называют дисперсионным отношением.

Здесь s_M^2 — межгрупповая дисперсия,

$$s_M^2 = \frac{1}{n} \sum_{i=1}^r n_i (\bar{x}_{(i)} - \bar{x})^2; \quad (1.5)$$

s_r^2 — внутригрупповая дисперсия,

$$s_r^2 = \frac{1}{n} \sum_{i=1}^r \sum_{j=1}^{n_i} (x_{(i)j} - \bar{x}_{(i)})^2 = \frac{1}{n} \sum_{i=1}^r (n_i - 1) \tilde{s}_{(i)}^2; \quad (1.6)$$

\bar{x} — общее выборочное среднее,

$$\bar{x} = \frac{1}{n} \sum_{i=1}^r \sum_{j=1}^{n_i} x_{(i)j}. \quad (1.7)$$

В случае справедливости гипотезы H_0 и числитель, и знаменатель дисперсионного отношения будут иметь (при достаточно большом объеме выборки) распределение χ^2 с $r - 1$ и $n - r$ степенями свободы, соответственно. Поэтому дисперсионное отношение в случае справедливости H_0 для достаточно больших объемов выборки будет иметь распределение Фишера — Снедекора с $r - 1, n - r$ степенями свободы.

Таким образом, по уровню значимости с помощью таблицы распределения Фишера — Снедекора можно определить критическое значение $F_{кр} = F_{1-\alpha; r-1, n-r}$ как квантиль уровня $1 - \alpha$ распределения Фишера — Снедекора с числом степеней свободы $k_1 = r - 1, k_2 = n - r$.

Далее, если значение статистики F меньше критического, то данные согласуются с гипотезой о равенстве средних и можно считать, что фактор A не влияет на результативный признак X .

Если значение статистики F больше критического, то данные противоречат гипотезе о равенстве средних и можно считать, что уровни фактора A по-разному влияют на результативный признак X .

Пример 1.9. Для данных, приведенных в примере 1.8, требуется проверить гипотезу дисперсионного анализа о равенстве средних (табл. 1.12).

Таблица 1.12

К примеру 1.9

Номер бригады i	Групповые средние $\bar{x}_{(i)}$	Групповые дисперсии $\tilde{S}_{(i)}^2$
1	142,75	4,92
2	150,25	1,58
3	147,50	1,67
4	152,75	4,92

Решение

Расчеты проводим по формулам (1.4)—(1.7).

Общее выборочное среднее $\bar{x} = \frac{1}{16} \sum_{i=1}^4 \sum_{j=1}^4 x_{(i)j} = 148,31$.

$$\text{Межгрупповая дисперсия } s_M^2 = \frac{1}{16} \sum_{i=1}^4 4(\bar{x}_{(i)} - 148,31)^2 = 13,762.$$

$$\text{Внутригрупповая дисперсия } s_r^2 = \frac{1}{16} \sum_{i=1}^r 3 \cdot \tilde{s}_{(i)}^2 = 2,454.$$

$$\text{Дисперсионное отношение: } F = \frac{s_M^2 / (r-1)}{s_r^2 / (n-r)} = \frac{12 \cdot 13,762}{3 \cdot 2,454} = 22,43.$$

По таблицам распределения Фишера — Снедекора для $\alpha = 0,05$ степеней свободы $k_1 = 3$, $k_2 = 12$ найдем критическое значение $F_{\text{кр}} = 3,49$.

Так как дисперсионное отношение больше критического, то гипотезу H_0 отклоняем и считаем, что объем ежедневной выработки зависит от работающей бригады.

Два последних рассмотренных критерия используют в важном разделе математической статистики — дисперсионном анализе.

Основные понятия дисперсионного анализа

В настоящее время дисперсионный анализ определяется как статистический метод, предназначенный для выявления влияния отдельных факторов на результат эксперимента.

Современные приложения дисперсионного анализа охватывают широкий круг задач экономики, социологии, биологии и техники.

Модели дисперсионного анализа в зависимости от числа факторов, влияние которых изучается, подразделяются на однофакторные, двухфакторные и т.д.

Изучаемую случайную величину Y (на которую влияют один или несколько факторов) в дисперсионном анализе называют результативным признаком.

Проиллюстрируем сказанное. Пусть директора фирмы интересуется зависимость выполнения работ за смену от работающей на стройке бригады. Всего на стройке работают r бригад. Объем выполненных работ является результативным признаком Y , работающая бригада — это фактор, который влияет на результативный признак, а уровень фактора — это конкретная бригада. Число уровней фактора r совпадает с числом работающих бригад. В данном случае рассматривается однофакторная детерминированная модель дисперсионного анализа.

Суть метода дисперсионного анализа состоит в разложении общей вариации изучаемого показателя на части, соответствующие раздельному и совместному влиянию факторов, и в статистическом изучении этих частей с целью выяснения приемлемости гипотез о существовании влияний факторов.

Рассмотрим подробно однофакторную модель дисперсионного анализа.

Пусть имеется результирующий признак Y и фактор X , с уровнями x_1, x_2, \dots, x_r . При i -м уровне фактора X (где i меняется от 1 до r) случайная величина Y принимает значения $y_{(i)1}, y_{(i)2}, \dots, y_{(i)n_i}$.

В основе дисперсионного анализа лежит следующая вероятностная модель:

$$y_{(i)j} = m + a_{(i)} + \varepsilon_{(i)j},$$

где m — математическое ожидание случайной величины Y , $m = M(Y)$;

$a_{(i)}$ — эффект влияния уровня i -го фактора на результирующий признак, $a_{(i)} = M(Y|x_i) - M(Y)$;

$\varepsilon_{(i)j}$ — случайный остаток, отражающий влияние всех остальных (неучтенных) факторов.

Основными предположениями дисперсионного анализа являются следующие:

- Остатки $\varepsilon_{(i)j}$ взаимно независимы для любых i и j .
- Распределение остатков $\varepsilon_{(i)j}$ является нормальным с нулевым средним $M(\varepsilon_{(i)j}) = 0$, постоянной дисперсией $D(\varepsilon_{(i)j}) = \sigma_0^2$ и не зависит от i и j .

Зависимость $a_{(i)}$ от уровня фактора i указывает на влияние фактора на результирующий признак, и, следовательно, если $a_{(i)}$ не зависят от уровня фактора i , то это говорит о том, что фактор X не влияет на результирующий признак Y .

Целью исследования является выяснение изменчивости величин $a_{(i)}$ в зависимости от уровня фактора. Если $a_1 = a_2 = \dots = a_r$, то можно сделать вывод о том, что фактор X не влияет на результирующий признак Y .

Основная идея дисперсионного анализа основана на разложении суммы квадратов отклонений.

Исходными данными в дисперсионном анализе являются r групп наблюдений над случайной величиной Y (группа характеризуется определенным значением фактора):

$$\begin{aligned}
 & Y_{(1)1}, Y_{(1)2}, \dots, Y_{(1)n_1}; \\
 & Y_{(2)1}, Y_{(2)2}, \dots, Y_{(2)n_2}; \\
 & \dots \\
 & Y_{(r)1}, Y_{(r)2}, \dots, Y_{(r)n_r}.
 \end{aligned}$$

Общее количество наблюдений (объем выборки):

$$n = \sum_{i=1}^r n_i.$$

Оценкой математического ожидания (согласно методу моментов) генеральной случайной величины Y является среднее арифметическое всех выборочных значений:

$$\bar{y} = \frac{1}{n} \sum_{i=1}^r \sum_{j=1}^{n_i} y_{(i)j}. \quad (1.8)$$

Оценкой дисперсии (по методу моментов) генеральной случайной величины X является нормированная сумма квадратов:

$$s^2 = \frac{1}{n} \sum_{i=1}^r \sum_{j=1}^{n_i} (y_{(i)j} - \bar{y})^2.$$

Основная идея дисперсионного анализа заключается в разбиении последней суммы квадратов отклонений на несколько компонент, каждая из которых соответствует предполагаемой причине изменения значений $a_{(i)} = M(Y|x_i) - M(Y)$.

Оценка условного математического ожидания $M(Y|x_i)$ — это среднее арифметическое выборочных данных в i -й группе:

$$\bar{y}_{(i)} = \frac{1}{n_i} \sum_{j=1}^{n_i} y_{(i)j}.$$

Оценка условной дисперсии $\sigma_{(i)}^2 = D(Y|x_i)$:

$$\tilde{\sigma}_{(i)}^2 = \frac{1}{n_i} \sum_{j=1}^{n_i} (y_{(i)j} - \bar{y}_{(i)})^2.$$

Можно показать, что

$$\frac{1}{n} \sum_{i=1}^r \sum_{j=1}^{n_i} (y_{(i)j} - \bar{y})^2 = \frac{1}{n} \sum_{i=1}^r \sum_{j=1}^{n_i} (y_{(i)j} - \bar{y}_{(i)})^2 + \frac{1}{n} \sum_{i=1}^r n_i (\bar{y}_{(i)} - \bar{y})^2.$$

В левой части написанного равенства стоит s^2 — оценка дисперсии генеральной случайной величины Y , в правой части — два слагаемых.

Второе слагаемое в правой части — это нормированная сумма квадратов между группами, которая является дисперсией групповых средних, т.е. это *межгрупповая дисперсия*:

$$\frac{1}{n} \sum_{i=1}^r n_i (\bar{y}_{(i)} - \bar{y})^2 = s_M^2.$$

Межгрупповая дисперсия отражает влияние фактора X на генеральную случайную величину Y .

Первое слагаемое в правой части, нормированная сумма квадратов внутри групп, представляет собой среднюю из групповых дисперсий, т.е. это *внутригрупповая дисперсия*:

$$\frac{1}{n} \sum_{i=1}^r \sum_{j=1}^{n_i} (y_{(i)j} - \bar{y}_{(i)})^2 = \frac{1}{n} \sum_{i=1}^r n_i \cdot \tilde{\sigma}_{(i)}^2 = s_r^2.$$

Внутригрупповая дисперсия отражает влияние остаточных факторов, не включенных в модель.

Итак, верно *правило сложения дисперсий*: $s^2 = s_r^2 + s_M^2$, т.е. *общая дисперсия наблюдений равна сумме внутригрупповой дисперсии и межгрупповой дисперсии*.

При решении практических задач с использованием модели однофакторного дисперсионного анализа нужно выполнить следующую последовательность действий:

- проверить условие применимости модели, т.е. равенство дисперсий в каждой группе (с помощью критерия Бартлетта);
- проверить гипотезу о влиянии уровней фактора на результативный признак (через проверку гипотезы о равенстве групповых средних);
- оценить степень влияния уровней фактора на результативный признак и оценить параметры модели.

Для измерения степени влияния фактора X на результативный признак Y используют выборочный коэффициент детерминации:

$$\bar{d} = \frac{s_M^2}{s^2} = \frac{s_M^2}{s_M^2 + s_\Gamma^2},$$

который показывает, какую долю выборочной дисперсии составляет дисперсия групповых средних. Иначе говоря, какая доля общей дисперсии объясняется зависимостью результативного признака Y от фактора X .

Далее, если уровни фактора X по-разному влияют на результативный признак Y (отвергнута гипотеза о равенстве средних), то оценка математического ожидания m генеральной случайной величины Y производится как среднее арифметическое всех выборочных данных по формуле (1.8).

Оценку величин $a_{(i)} = M(Y|x_i) - M(Y)$ (влияния i -го уровня фактора X) производят по формуле

$$\bar{a}_{(i)} = \bar{y}_{(i)} - \bar{y}, \text{ для } i = 1, 2, \dots, r,$$

где $\bar{y}_{(i)}$ — оценка условного математического ожидания, которая представляет собой выборочное среднее по i -й группе наблюдений;

$$\bar{y}_{(i)} = \frac{1}{n_i} \sum_{j=1}^{n_i} y_{(i)j}.$$

Пример 1.10. На основе данных примеров 1.8, 1.9 требуется оценить степень зависимости объема ежедневной выработки работающей бригады с помощью коэффициента детерминации.

Решение

Было получено, что $s_M^2 = 13,762$, $s_\Gamma^2 = 2,454$, следовательно, коэффициент детерминации

$$\bar{d} = \frac{s_M^2}{s_M^2 + s_\Gamma^2} = \frac{13,762}{13,762 + 2,454} = 0,849.$$

Это означает, что 84,9% общей вариации ежедневного объема выработки связано с работающей сменой.

Оценка математического ожидания (средней выработки на стройке) $m = M(Y)$ равна $\bar{y} = 148,31$.

Оценки параметров $a_{(i)} = M(Y|x_i) - M(Y)$ соответственно равны:

$$\tilde{a}_{(1)} = \bar{y}_{(1)} - \bar{y} = 142,75 - 148,31 = -5,56;$$

$$\tilde{a}_{(2)} = \bar{y}_{(2)} - \bar{y} = 150,15 - 148,31 = 1,96;$$

$$\tilde{a}_{(4)} = \bar{y}_{(4)} - \bar{y} = 152,75 - 148,31 = 4,44.$$

1.4.5. ПРОВЕРКА ГИПОТЕЗ О ВИДЕ РАСПРЕДЕЛЕНИЯ

При проверке статистических гипотез не всегда есть основание высказать альтернативную гипотезу в явном виде. Часто имеется в виду просто невыполнение основной гипотезы. В этом случае требуется выяснить, согласуются ли данные с основной гипотезой или противоречат ей. Критерии, которые позволяют это сделать, называют *критериями согласия*.

Анализируя выборочные данные — эмпирические частоты и гистограмму (или полигон), можно высказать предположение о виде закона распределения генеральной случайной величины X .

При этом ввиду ограниченного числа наблюдений, опытный (эмпирический) закон распределения будет в какой-то мере отличаться от предполагаемого, даже если предположение о законе распределения сделано правильно. В связи с этим возникает необходимость решать следующую задачу: является ли расхождение между опытным и предполагаемым законами распределения следствием ограниченного числа опытов или оно является существенным, и истинное распределение отличается от предполагаемого.

Для окончательного решения вопроса о виде закона распределения в подобных случаях представляется целесообразным проверить, насколько сделанное предположение согласуется с опытом. Так как все предположения о виде теоретического распределения — гипотезы, а не категорические утверждения, то они, естественно, должны быть подвергнуты статистической проверке с помощью критериев согласия относительно закона распределения.

Критерии согласия, опирающиеся на установленный закон распределения, дают возможность установить, когда расхождение между эмпирическими и теоретическими частотами следует признать несущественным (случайным), а когда — существенным (неслучайным). Таким образом, критерии согласия позволяют отвергнуть или подтвердить правильность выдвинутой гипотезы о виде

распределения в эмпирическом ряду и дать ответ, можно ли принять для данного эмпирического распределения модель, выраженную некоторым теоретическим законом распределения.

Критерий согласия χ^2 (хи-квадрат) Пирсона является одним из основных статистических критериев проверки гипотез о виде распределения.

Основная гипотеза состоит в том, что выборка произведена из совокупности, имеющей некоторое заданное распределение. Основная гипотеза может быть простой (т.е. функция распределения определена однозначно) или параметрической (функция распределения зависит от k неизвестных параметров).

Процедура проверки гипотезы о виде распределения с помощью критерия χ^2 Пирсона состоит из пяти этапов:

- Весь диапазон значений генеральной случайной величины X разбивают на интервалы $\Delta_1, \Delta_2, \dots, \Delta_k$ без общих точек и подсчитывают число наблюдений, попавших в каждый интервал (интервалы выбирают так, чтобы в каждый попало не менее пяти наблюдений).
- Предположив справедливость основной гипотезы, подсчитывают вероятность попадания в каждый интервал:

$$p_i = P\{X \in \Delta_i | H_0\}; i = 1, 2, \dots, k.$$

При проверке параметрической гипотезы о виде распределения параметры распределения оказываются неизвестными, поэтому их заменяют наилучшими оценками по имеющейся выборке.

- Составляют статистику критерия:

$$\chi^2(k-l-1) = \sum_{i=1}^k \frac{(n_i - n \cdot p_i)^2}{n \cdot p_i},$$

где l — число неизвестных параметров распределения.

Критическая область — это область больших значений статистики критерия. Критическое значение $\chi_{кр}^2$ — квантиль уровня $1 - q$ распределения χ^2 с $k - l - 1$ степенью свободы.

Если значение статистики критерия меньше критического значения, то гипотезу принимают. В этом случае говорят, что выборочные данные согласуются с гипотезой или не противоречат ей. Если значение статистики критерия больше (или равно) критическому значению, то гипотезу отвергают. В этом случае говорят, что

выборочные данные не согласуются с гипотезой или противоречат ей.

Приведем таблицу для определения критических значений критерия χ^2 Пирсона при заданном уровне значимости (табл. 1.13).

Таблица 1.13

Значения критерия Пирсона

Уровень значимости	Число степеней свободы								
	1	2	3	4	5	7	9	11	15
0,05	3,84	5,99	7,82	9,49	11,1	14,1	16,9	19,7	25,0
0,01	6,64	9,21	11,3	13,3	15,1	18,5	21,7	24,7	30,6

При использовании данной методики требуется, чтобы объем выборки $n > 50$.

Пример 1.11. Через равные промежутки времени в тонком растворе золота регистрировалось количество частиц золота, попавших в поле зрения микроскопа (табл. 1.14).

Таблица 1.14

К примеру 1.11

Число частиц	Частоты n_i	Число частиц	Частоты n_i
0	112	5	5
1	168	6	1
2	130	7	1
3	68	Всего	518
4	32	—	—

Требуется проверить, что количество частиц золота, попавших в поле зрения микроскопа, имеет распределение Пуассона.

Решение

Распределение Пуассона — это дискретное однопараметрическое распределение, вероятности которого определены формулами

$$p_j = P(X = j) = e^{-\lambda} \frac{\lambda^j}{j!} \text{ для } j = 0; 1; 2; \dots$$

Сначала оценим параметр распределения Пуассона:

$$\hat{\lambda} = \bar{X} = \frac{1}{518} (0 \cdot 112 + 1 \cdot 168 + 2 \cdot 130 + 3 \cdot 68 + 4 \cdot 32 + 5 \cdot 5 + 6 \cdot 1 + 7 \cdot 1) \approx 1,54.$$

Далее суммируем три последних значения (5 + 1 + 1) табл. 1.14 и сведем их в табл. 1.15.

Таблица 1.15

К примеру 1.11

Число частиц	Частоты n_i	Число частиц	Частоты n_i
0	112	4	32
1	168	5 и более	7
2	130	Всего	518
3	68	—	—

Теперь вычислим оценки для теоретических вероятностей:

$$\hat{p}_0 = e^{-\hat{\lambda}} = \exp(-1,54); \quad p_1 = \hat{\lambda} \cdot e^{-\hat{\lambda}} = 1,54 \cdot \exp(-1,54);$$

$$\hat{p}_4 = \frac{\hat{\lambda}^4}{4!} e^{-\hat{\lambda}} = \frac{(1,54)^4}{24} \exp(-1,54); \quad \hat{p}_5 = 1 - \sum_{i=1}^4 \hat{p}_i.$$

Количество степеней свободы: $k - 2 = 6 - 2 = 4$.

Значение статистики критерия: $\chi^2(4) = 7,85$

Критическое значение найдем по табл. П3 приложения:
 $\chi_{кр}^2 = 9,49$.

Согласие имеет место.

Нормально распределенная случайная величина X имеет плотность распределения

$$f(x) = \frac{1}{\sqrt{2\pi} \cdot \sigma} \exp\left\{-\frac{(x-m)^2}{2\sigma^2}\right\}.$$

Эта плотность зависит от двух параметров: m — математического ожидания случайной величины X и σ^2 — дисперсии случайной величины X .

Оценки параметров нормального распределения следующие:

$$\tilde{m} \text{ — выборочное среднее, } \tilde{m} = \bar{x} = \frac{1}{n} \sum_j x_{(j)} n_j;$$

$$\tilde{\sigma}^2 \text{ — «подправленная» выборочная дисперсия, } \tilde{\sigma}^2 = \frac{1}{n-1} \sum_j (x_{(j)} - \bar{x})^2 n_j.$$

Здесь $x_{(j)}$ — середина j -го интервала группировки; n_j — число элементов выборки, попавших в j -й интервал группировки.

Теоретическая вероятность попадания в j -й интервал группировки:

$$p_j = P\{X \in \Delta_j \mid H_0\};$$

$$p_j = \int_{x \in \Delta_j} f(x) dx.$$

В частности, если произведено разбиение на одинаковые по длине интервалы $\Delta_j = (x_{(j)} - \delta; x_{(j)} + \delta)$ длины 2δ , то

$$p_j = F(x_{(j)} + \delta) - F(x_{(j)} - \delta) = F_{N(0;1)}\left(\frac{x_{(j)} + \delta - \tilde{m}}{\tilde{\sigma}}\right) - F_{N(0;1)}\left(\frac{x_{(j)} - \delta - \tilde{m}}{\tilde{\sigma}}\right).$$

Здесь $F_{N(0;1)}$ — функция распределения стандартно нормально-го закона $F_{N(0;1)}(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt$ (ее значения находят по соответствующим таблицам).

Для того чтобы упростить вычисления, можно пользоваться приближенной формулой

$$\begin{aligned} P(\xi \in [a; b]) &= F_{N(0;1)}(b) - F_{N(0;1)}(a) \approx |b - a| f_{N(0;1)}\left(\frac{a+b}{2}\right) = \\ &= \frac{|b-a|}{\sqrt{2\pi}} \exp\left(-\frac{(a+b)^2}{8}\right), \end{aligned} \quad (1.9)$$

где ξ — случайная величина, имеющая стандартно нормальное распределение.

Применяя данную формулу, получаем, что теоретическую вероятность попадания в j -й интервал группировки $\Delta_j = (x_{(j)} - \delta; x_{(j)} + \delta)$ можно вычислить без применения таблиц значений функции распределения стандартно нормального закона:

$$p_j \approx \frac{\delta}{\tilde{\sigma}} \sqrt{\frac{2}{\pi}} \exp\left(-\frac{(x_{(j)} - \tilde{m})^2}{2\tilde{\sigma}^2}\right).$$

Если гипотеза о том, что распределение генеральной случайной величины является нормальным, проверяется с помощью критерия Пирсона, то статистика критерия имеет вид

$$\chi^2(k-3) = \sum_{j=1}^m \frac{(n_j - n \cdot p_j)^2}{n \cdot p_j}.$$

Пример 1.12. Имеются следующие данные распределения 200 проб нити по крепости (табл. 1.16).

Таблица 1.16

К примеру 1.12

Крепость нити	Число проб	Крепость нити	Число проб
120—130	1	160—170	56
130—140	8	170—180	34
140—150	27	180—190	14
150—160	58	190—200	2

Требуется проверить гипотезу о том, что крепость нити — нормально распределенная случайная величина.

Решение

Сначала оценим параметры распределения:

$$\tilde{m} = \bar{x} = \frac{1}{n} \sum_j x_{(j)} n_j = 161,4;$$

$$\tilde{\sigma}^2 = \frac{1}{n-1} \sum_j (x_{(j)} - \bar{x})^2 n_j = 13^2.$$

Дальнейшие вычисления представим в виде табл. 1.17.

Таблица 1.17

К примеру 1.12

Крепость нити	n_j	$\frac{x_{(j)} - \bar{x}}{\tilde{\sigma}}$	$n_j p_j$	Крепость нити	n_j	$\frac{x_{(j)} - \bar{x}}{\tilde{\sigma}}$	$n_j p_j$
120—130	1	-2,80	1	160—170	56	0,28	59
130—140	8	-2,03	8	170—180	34	1,05	35
140—150	27	-1,26	28	180—190	14	1,82	12
150—160	58	-0,49	55	190—200	2	2,58	2

Примечание. Здесь $x_{(j)}$ — середина j -го интервала группировки; n_j — число элементов выборки, попавших в j -й интервал группировки.

Вероятность попадания в j -й интервал группировки рассчитывается по формуле

$$p_j = \frac{5}{13} \sqrt{\frac{2}{\pi}} \exp\left(-\frac{(x_{(j)} - 161,4)^2}{338}\right).$$

Как видно из таблицы, теоретические частоты близки к эмпирическим, тем не менее отдельные расхождения имеют место. Для случайности или существенности этих расхождений используем критерии согласия.

Статистика критерия Пирсона $\chi^2(8 - 3) = 0,71$.

Критическое значение найдем по табл. ПЗ приложения: $\chi_{0,95; 5}^2 = 11,7$.

Следовательно, значение статистики значительно меньше критического, согласие хорошее.

1.5. Корреляционно-регрессионный анализ

1.5.1. Понятие корреляционного анализа

Корреляционный анализ — раздел математической статистики, в котором изучаются задачи выявления статистических зависимостей между случайными величинами путем оценок различных коэффициентов корреляции. Методы корреляционного анализа дают хорошие результаты тогда, когда данные эксперимента можно считать выбранными из генеральной совокупности, распределенной по многомерному нормальному закону.

Корреляционной зависимостью между случайными величинами X и Y называется функциональная зависимость между значениями одной из них и условным математическим ожиданием другой.

На координатной плоскости изобразим все выборочные точки $(x_1; y_1), (x_2; y_2), \dots, (x_n; y_n)$ по выборке корреляционной зависимости случайных величин X и Y . Это изображение называется *корреляционным полем* (рис. 1.14).

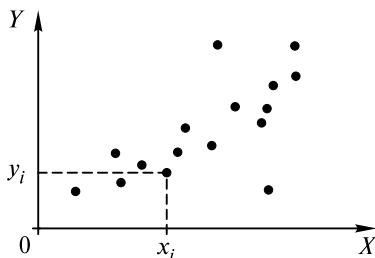


Рис. 1.14. Корреляционное поле

Свойства коэффициента корреляции следующие:

- Величина коэффициента корреляции заключена в пределах $[-1; 1]$.

- Для независимых случайных величин коэффициент корреляции равен 0.

Случайные величины, имеющие нулевой коэффициент корреляции, называют *некоррелированными*.

Из независимости случайных величин следует их некоррелированность. Обратное неверно. Условие независимости более жесткое, чем условие некоррелированности.

Для непрерывных распределений в качестве x_i и y_i выбираются середины интервалов группировки.

Коэффициент корреляции между случайными переменными X и Y определяется как

$$\rho = \frac{M(X \cdot Y) - M(X) \cdot M(Y)}{\sigma(X) \cdot \sigma(Y)}.$$

- Коэффициент корреляции характеризует линейную зависимость между случайными величинами. Если случайные величины связаны линейной зависимостью $Y = a \cdot X + b$, то коэффициент корреляции по модулю равен 1. Можно показать, что обратное утверждение также верно.

- Если коэффициент корреляции больше нуля, то говорят о положительной корреляции между случайными величинами, это означает, что при возрастании одной из случайных величин другая имеет тенденцию в среднем возрастать. Например, рост и вес человека связаны положительной корреляцией.

- Если коэффициент корреляции меньше нуля, то говорят об отрицательной корреляции между случайными величинами, это означает, что при возрастании одной из случайных величин другая имеет тенденцию в среднем убывать.

Так, время, потраченное на регулировку прибора, и количество неисправностей, обнаруженных при работе прибора, связаны отрицательной корреляцией.

Оценкой коэффициента корреляции является *выборочный коэффициент корреляции*, который можно вычислить как

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{(n-1)s_x s_y},$$

где \bar{x} — выборочное среднее для величины X ;

\bar{y} — выборочное среднее для величины Y ;

s_x — выборочное среднеквадратическое для величины X ;

s_y — выборочное среднеквадратическое для величины Y .

Формулу для выборочного коэффициента корреляции можно преобразовать:

$$r = \frac{n \sum_{i=1}^n x_i y_i - \left(\sum_{i=1}^n x_i \right) \left(\sum_{i=1}^n y_i \right)}{\sqrt{\left[n \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i \right)^2 \right] \left[n \sum_{i=1}^n y_i^2 - \left(\sum_{i=1}^n y_i \right)^2 \right]}}. \quad (1.10)$$

После вычисления выборочного коэффициента корреляции проверяют гипотезу о значимости связи, так как выборочный коэффициент корреляции может не равняться нулю из-за отбора переменных в выборку.

Обычно проверяется основная гипотеза об отсутствии корреляционной связи:

$H_0: r = 0$ против альтернативы $H_1: r \neq 0$.

Для проверки используют статистику Стьюдента:

$$St = \frac{r \cdot \sqrt{n-2}}{\sqrt{1-r^2}}. \quad (1.11)$$

Считают, что выборочный коэффициент корреляции значимо отличается от нуля, если значение статистики по модулю больше критического:

$$|St| > t_{1-q, n-2},$$

где $t_{1-q, n-2}$ — критическое значение распределения Стьюдента, определенное на уровне значимости $1 - q$ при числе степеней свободы, равном $n - 2$ (т.е. квантиль уровня $q/2$ распределения Стьюдента с $n - 2$ степенями свободы).

Пример 1.13. На прессе определяли прочность образцов в возрасте 28 суток в зависимости от времени изотермической выдержки.

Требуется проверить, можно ли считать значимым коэффициент корреляции между прочностью и временем изотермической обработки.

Решение

Каждый опыт проводился на трех образцах. Результаты эксперимента занесены в табл. 1.18.

Таблица 1.18

Результаты эксперимента

№ п/п	Время изотермической выдержки, ч	№ образца	Прочность образцов, МПа
1	3	1	22,3
		2	21,3
		3	21,6
2	4	1	21,8
		2	22,1
		3	20,3
3	5	1	21,5
		2	20,3
		3	20,8
4	6	1	20,9
		2	20,8
		3	21,5
5	7	1	19,8
		2	20,8
		3	20,9

Диаграмма рассеивания в данном случае имеет вид, приведенный на рис. 1.15.

Найдем выборочный коэффициент корреляции по формуле (1.10):

$$r = \frac{15 \sum_{i=1}^{15} x_i y_i - \left(\sum_{i=1}^{15} x_i \right) \left(\sum_{i=1}^{15} y_i \right)}{\sqrt{\left[15 \sum_{i=1}^{15} x_i^2 - \left(\sum_{i=1}^{15} x_i \right)^2 \right] \left[15 \sum_{i=1}^{15} y_i^2 - \left(\sum_{i=1}^{15} y_i \right)^2 \right]}} = -0,58.$$

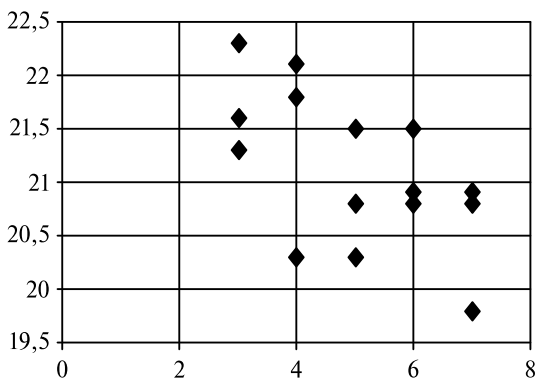


Рис. 1.15. Диаграмма рассеивания

Проверим гипотезу об отсутствии корреляционной связи, для этого вычислим статистику Стьюдента по (1.11):

$$St = \frac{r \sqrt{n-2}}{\sqrt{1-r^2}} = \frac{-0,28 \sqrt{15-2}}{\sqrt{1-(-0,28)^2}} = -2,6.$$

По табл. П1 приложения найдем $t_{1-q, n-2}$ — критическое значение распределения Стьюдента, задав уровень значимости 0,05 (число степеней свободы равно 13): $t_{0,05; 13} = 2,16$.

Значение статистики меньше критического, поэтому считаем, что коэффициент корреляции значимо отличается от нуля.

1.5.2. МНОЖЕСТВЕННАЯ КОРРЕЛЯЦИЯ

Множество задач экономики, производства, техники описывается многофакторными моделями, т.е. корреляционные связи не ограничиваются связями между двумя признаками — результативным (Y) и факторным (X).

Результативный признак, как правило, зависит от нескольких факторов. Например, инфляция тесно связана с динамикой цен, объемом производства, численностью безработных и другими факторами.

В условиях действия множества факторов показатели парной корреляции оказываются условными и неточными. В связи с этим встает вопрос о выяснении тесноты связи результативного признака с различными факторами.

Рассмотрим совокупность k случайных величин X_1, X_2, \dots, X_k . Составим матрицу выборочных парных коэффициентов корреляции:

$$\mathfrak{R} = \begin{pmatrix} 1 & r_{12} & \dots & r_{1k} \\ r_{21} & 1 & \dots & r_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ r_{k1} & r_{k2} & \dots & 1 \end{pmatrix}.$$

Теснота линейной связи случайной величины X_i с совокупностью остальных переменных измеряется с помощью *множественного коэффициента корреляции*.

Выборочный множественный коэффициент корреляции определяется по формуле

$$R_i = \sqrt{1 - \frac{\det(\mathfrak{R})}{\mathfrak{R}_{ii}}},$$

где $\det(\mathfrak{R})$ — определитель матрицы парных коэффициентов корреляции; \mathfrak{R}_{ii} — алгебраическое дополнение элемента, стоящего на пересечении i -й строки и i -го столбца.

В частности, для трех случайных величин X, Y и Z можно получить (положив $k = 3$), что множественный коэффициент корреляции равен

$$R_Z = \sqrt{\frac{r_{Z,X}^2 + r_{Z,Y}^2 - 2r_{Z,X}r_{Z,Y}r_{X,Y}}{1 - r_{X,Y}^2}}. \quad (1.12)$$

Множественный коэффициент корреляции обладает двумя важными свойствами:

- $0 \leq R_i \leq 1$ для любого i (чем ближе коэффициент к 1, тем ближе связь i -й переменной с остальными коэффициентами к линейной);
- $|r_{ij}| \leq R_i$ для любых i и j .

Величина R_i^2 , называемая *выборочным множественным коэффициентом детерминации*, показывает, какой вклад в дисперсию i -й случайной величины вносят остальные величины.

Наряду с измерением совместного влияния отобранных факторов на результативный признак важно определить воздействие каж-

дого фактора при изолированном рассмотрении его взаимосвязи с результативным признаком.

Это возможно, когда остальные факторы зафиксированы на постоянном уровне.

Для решения этой задачи применяют *частные коэффициенты корреляции*

$$R_{ij} = \frac{-\mathfrak{R}_{ij}}{\sqrt{\mathfrak{R}_{ii}\mathfrak{R}_{jj}}},$$

где \mathfrak{R}_{ij} — алгебраическое дополнение стоящего на пересечении i -й строки и j -го столбца элемента матрицы выборочных парных коэффициентов корреляции.

Частный коэффициент корреляции по своим свойствам не отличается от парного.

В частности, для трех случайных величин X , Y и Z можно получить (положив $k = 3$), что частный коэффициент корреляции имеет вид

$$R_{Z,X} = \frac{r_{Z,X} - r_{Z,Y} r_{X,Y}}{\sqrt{(1 - r_{Z,Y}^2)(1 - r_{X,Y}^2)}}. \quad (1.13)$$

Пример 1.14. В табл. 1.19 приведены значения выборочных парных коэффициентов корреляции, характеризующих уровень доходов работающего населения России.

Требуется рассчитать множественный и частные коэффициенты корреляции и выборочный множественный коэффициент детерминации.

Таблица 1.19

Коэффициенты корреляции к примеру 1.14

Показатели	Среднедневная заработная плата X , р.	Средний возраст Y , годы	Среднедушевой доход Z , р.
Среднедневная заработная плата X , р.	1	-0,1160	0,8405
Средний возраст безработного Y , годы	-0,1160	1	-0,2101
Среднедушевой доход Z , р.	0,8405	-0,2101	1

Решение

По формуле (1.13) рассчитываем частные коэффициенты корреляции:

$$R_{Z,X} = 0,8404; R_{Z,Y} = -0,2092; R_{X,Y} = 0,1144.$$

Множественный коэффициент корреляции рассчитываем по формуле (1.12): $R_Z = 0,8481$, коэффициент детерминации равен $R_Z^2 = 0,7193 \approx 0,72$.

Из-за слабой связи между X и Y коэффициенты парной и частной корреляции отличаются незначительно.

Зависимость Z от X и Y характеризуется как тесная, в которой 72 % изменения среднего душевого дохода определяются изменением учтенных факторов: средней заработной платы и среднего возраста безработного.

1.5.3. ВВЕДЕНИЕ В РЕГРЕССИОННЫЙ АНАЛИЗ

Регрессионная зависимость является частным случаем стохастической зависимости и подразумевает зависимость среднего значения величины Y от другой случайной величины X (одномерной или многомерной).

Регрессионная зависимость Y от X проявляется в изменении средних значений Y при изменении X , хотя при каждом фиксированном значении $X = x$ величина Y остается случайной величиной с определенным распределением.

В математической статистике имеют дело с оценками соответствующих вероятностных характеристик, поэтому в качестве оценки условного математического ожидания принимают условное среднее.

Если при каждом значении $x = x_i$ наблюдается n_i значений $y_1^{(i)}$, $y_2^{(i)}$, ..., $y_{n_i}^{(i)}$ величины y , то зависимость средних арифметических

$$\bar{y}^{(i)} = \frac{y_1^{(i)} + y_2^{(i)} + \dots + y_{n_i}^{(i)}}{n_i}$$

от x_i и является регрессией в статистическом понимании этого термина.

Если число наблюдений, соответствующее некоторым значениям X , недостаточно велико, то такой метод может привести к ненадежным результатам.

Уравнение $y = y(x)$, в котором x играет роль «независимой» переменной, называют *уравнением регрессии*, а соответствующий график — *линией* или *кривой регрессии*.

Линия регрессии может быть приближенно восстановлена по достаточно обширной корреляционной таблице: за приближенное значение $y(x)$ принимают среднее из тех выявленных значений Y , которым соответствует значение $X = x$.

Для выяснения вопроса, насколько хорошо регрессия передает изменение Y при изменении X , используется условная дисперсия Y при данном значении $X = x$ — *дисперсия Y относительно линии регрессии* (мера рассеивания относительно линии регрессии):

$$\sigma^2(x) = M[Y - M(Y|X=x)]^2 = M[Y - y(x)]^2.$$

При точной функциональной зависимости величина Y при данном $X = x$ принимает лишь одно определенное значение, т.е. рассеивание вокруг линии регрессии равно нулю. Наиболее простым является тот случай, когда регрессия Y по X линейна:

$$y(x) = b_0 + b_1 \cdot x,$$

где числа b_0 и b_1 называют коэффициентами регрессии.

Коэффициенты линейной регрессии вычисляют по формулам

$$b_0 = m_Y - r \frac{\sigma_Y}{\sigma_X} m_X, \quad b_1 = r \frac{\sigma_Y}{\sigma_X}.$$

Здесь $m_X = M(X)$, $m_Y = M(Y)$, $\sigma_X^2 = D(X)$, $\sigma_Y^2 = D(Y)$, $r = \frac{\text{cov}(X, Y)}{\sigma_X \sigma_Y}$.

Если двумерное распределение Y и X нормально, то линия регрессии Y по X (так же как и X по Y) является прямой с уравнением

$$y = m_Y + r \frac{\sigma_Y}{\sigma_X} (x - m_X).$$

В этом случае корреляционное отношение совпадает с коэффициентом корреляции и условная дисперсия не зависит от x (т.е. является постоянной величиной):

$$\sigma^2(x) = \sigma_Y^2 (1 - r^2).$$

Следовательно, коэффициент корреляции полностью определяет степень концентрации распределения вблизи линии регрессии.

Если регрессия Y по X отлична от линейной, то уравнение (1.9) является *линейным приближением* истинного уравнения регрессии.

Коэффициенты регрессии обычно неизвестны, и их оценивают по выборочным данным:

$$\hat{b}_1 = \frac{\overline{x \cdot y} - \bar{x} \cdot \bar{y}}{x^2 - \bar{x}^2}; \quad \hat{b}_0 = \bar{y} - \hat{b}_1 \cdot \bar{x}. \quad (1.14)$$

Линейная функция

$$\hat{y}(x) = \hat{b}_0 + \hat{b}_1 \cdot x$$

определяет *эмпирическую линию регрессии*, которая служит статистической оценкой неизвестной истинной линии регрессии.

Рассеивание вокруг линии регрессии можно оценить, используя эмпирическую среднюю дисперсию относительно линии регрессии:

$$\hat{\sigma}^2 = \frac{1}{n-2} \sum_{k=1}^n [y_k - \hat{y}(x_k)]^2.$$

Этот метод в предположении нормальной распределенности результатов наблюдений дает, в некотором смысле, оптимальные результаты и позволяет проводить экстраполяцию (прогнозирование) значений величины Y по имеющимся значениям величины X .

Оценку качества построенной модели дает средняя ошибка аппроксимации

$$\bar{A} = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}(x_i)}{y_i} \right| 100 \%. \quad (1.15)$$

Допустимый предел значений средней ошибки аппроксимации — не более 10 %.

Пример 1.15. Используем данные примера 1.13, в котором на прессе определяли прочность образцов в возрасте 28 суток в зависимости от времени изотермической выдержки.

Требуется рассчитать коэффициенты линейной регрессии прочности образцов.

Решение

Каждый опыт проводился на трех образцах. Результаты эксперимента занесены в табл. 1.19.

Ранее было доказано, что коэффициент корреляции между прочностью и временем изотермической выдержки значительно отличается от нуля. Поэтому получим линейную модель связи между прочностью и временем изотермической выдержки.

Рассчитаем коэффициенты линейной регрессии по формулам (1.14), для этого предварительно найдем: $\bar{x} \cdot \bar{y} = 105$; $\bar{y} = 21,1$; $\bar{x} = 5$; $\bar{x}^2 = 27$. Тогда

$$\hat{b}_1 = \frac{\bar{x} \cdot \bar{y} - \bar{x} \cdot \bar{y}}{\bar{x}^2 - \bar{x}^2} = \frac{105 - 21,1 \cdot 5}{27 - 25} = -0,25;$$

$$\hat{b}_0 = \bar{y} - \hat{b}_1 \cdot \bar{x} = 21,1 + 0,25 \cdot 5 = 22,75.$$

Таким образом, уравнение линейной регрессии имеет вид

$$\hat{y} = 22,75 - 0,25x$$

Рассеивание вокруг линии регрессии (рис. 1.16):

$$\hat{\sigma}^2 = \frac{1}{n-2} \sum_{k=1}^n [y_k - \hat{y}(x_k)]^2 = 0,52.$$

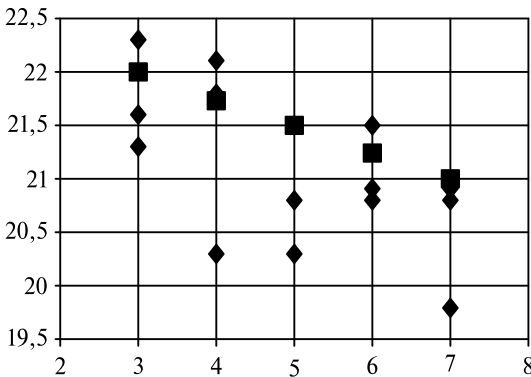


Рис. 1.16. Диаграмма рассеивания:

◆ — точки рассеивания; ■ — значения, полученные с помощью линейной регрессии

Оценим качества построенной модели:

$$\bar{A} = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}(x_i)}{y_i} \right| 100 \% = 2,5 \%$$

Средняя ошибка аппроксимации меньше допустимого предела на 10 %.

1.5.4. НЕЛИНЕЙНАЯ РЕГРЕССИЯ

В общем случае уравнение регрессии не является линейным.

Найти уравнение регрессии — значит по эмпирическим данным математически описать изменения зависимых случайных величин.

При выборе линии регрессии следует учитывать вид корреляционного поля и природу изучаемых величин со спецификой их взаимосвязей.

Для аналитической связи между случайными величинами часто используют следующие простые виды зависимостей:

$\hat{y}(x) = b_0 + b_1 \cdot x$ — линейную;

$\hat{y}(x) = b_0 + b_1 \cdot x + b_2 \cdot x^2$ — квадратичную;

$\hat{y}(x) = b_0 + b_1 \cdot \frac{1}{x}$ — гиперболическую;

$\hat{y}(x) = b_0 \cdot b_1^x$ — показательную;

$\hat{y}(x) = b_0 + b_1 \cdot \ln x$ — логарифмическую.

Выбрав тип функции, по эмпирическим данным определяют параметры уравнения. Для этой цели существует несколько методов. Наиболее часто используют метод наименьших квадратов. Его суть заключается в следующем требовании: искомые теоретические значения $\hat{y}(x_i)$ должны быть такими, при которых обеспечивалась бы минимальная сумма квадратов их отклонений от эмпирических значений y_i (сумма берется по всем проведенным опытам), т.е.

$$\sum_{i=1}^n [y_i - \hat{y}(x_i)]^2 \xrightarrow{b_1, b_2, \dots} \min.$$

Для нелинейной модели, так же как и для линейной, вычисляется средняя ошибка аппроксимации (1.15).

Более подробно о нелинейной регрессии будет рассказано в главе 2.

1.5.5. МНОЖЕСТВЕННАЯ РЕГРЕССИЯ

Методы множественного анализа позволяют решать задачу исследования зависимости одной переменной Y от нескольких переменных X_1, X_2, \dots, X_k . Для построения уравнения множественной регрессии чаще используют функции:

$$\hat{y}(x) = b_0 + b_1 \cdot x_1 + b_2 \cdot x_2 + \dots + b_k \cdot x_k \text{ — линейную;}$$

$$\hat{y}(x) = 1/b_0 + b_1 \cdot x_1 + b_2 \cdot x_2 + \dots + b_k \cdot x_k \text{ — гиперболическую;}$$

$$\hat{y}(x) = b_0 \cdot x_1^{b_1} \cdot x_2^{b_2} \cdot x_k^{b_k} \text{ — степенную;}$$

$$\hat{y}(x) = \exp(b_0 + b_1 \cdot x_1 + b_2 \cdot x_2 + \dots + b_k \cdot x_k) \text{ — экспоненту.}$$

Можно использовать и другие функции, приводимые к линейному виду.

Выбрать форму связи между переменными довольно сложно. Эта задача на практике основывается на априорном теоретическом анализе изучаемого явления.

Для оценки параметров уравнения множественной регрессии применяют метод наименьших квадратов.

Часто связь между результирующим признаком и анализируемыми факторами нелинейная, но она может быть сведена к линейной путем линеаризации с помощью замены переменной, поэтому рассмотрим более подробно линейное уравнение множественной регрессии:

$$\hat{y}(x) = b_0 + b_1 \cdot x_1 + b_2 \cdot x_2 + \dots + b_k \cdot x_k.$$

Следующая система нормальных уравнений, которые получены методом наименьших квадратов, позволяет найти оценки параметров $b_0, b_1, b_2, \dots, b_k$ линейного уравнения:

$$\left\{ \begin{array}{l} b_0 n + b_1 \sum_{i=1}^n x_{i1} + \dots + b_k \sum_{i=1}^n x_{ik} = \sum_{i=1}^n y_i, \\ b_0 \sum_{i=1}^n x_{i1} + b_1 \sum_{i=1}^n x_{i1}^2 + \dots + b_k \sum_{i=1}^n x_{i1} x_{ik} = \sum_{i=1}^n y_i x_{i1}, \\ \dots \\ b_0 \sum_{i=1}^n x_{ik} + b_1 \sum_{i=1}^n x_{i1} x_{ik} + \dots + b_k \sum_{i=1}^n x_{ik}^2 = \sum_{i=1}^n y_i x_{ik}. \end{array} \right.$$

Если ввести в рассмотрение матрицы:

$$X = \begin{pmatrix} 1 & x_{11} & \dots & x_{1k} \\ 1 & x_{21} & \dots & x_{2k} \\ \dots & \dots & \dots & \dots \\ 1 & x_{n1} & \dots & x_{nk} \end{pmatrix} \quad Y = \begin{pmatrix} y_1 \\ \dots \\ y_n \end{pmatrix} \quad B = \begin{pmatrix} b_1 \\ \dots \\ b_k \end{pmatrix}, \text{ то систему нормальных}$$

уравнений можно записать в матричном виде: $X^T \cdot X \cdot B = X^T \cdot Y$. Решением последней системы является вектор-столбец: $B = (X^T \cdot X)^{-1} \cdot X^T \cdot Y$.

Для того чтобы установить, соответствует ли выбранная регрессионная модель экспериментальным данным, используют критерий Фишера. По заданному уровню значимости q находят критическое значение распределения Фишера $F_{кр}$ при числе степеней свободы $k_1 = k - 1; k_2 = n - k$. Если значение статистики

$$F = \frac{(n-k) \sum_{i=1}^n [\hat{y}(x_i) - \bar{y}]^2}{(k-1) \sum_{i=1}^n [y_i - \hat{y}(x_i)]^2}$$

больше $F_{кр}$, то уравнение считают значимым (т.е. соответствующим экспериментальным данным на уровне q).

При этом выборочная остаточная дисперсия (с ее помощью оценивают неучтенные в модели случайные факторы) будет равна

$$s_0^2 = \frac{1}{n-k-1} \sum_{j=1}^n [y_j - \hat{y}(x_j)]^2.$$

Среднеквадратическое отклонение коэффициента регрессии равно $s(b_i) = s_0 \sqrt{d_{ii}}$ (здесь d_{ii} — диагональный элемент матрицы $(X^T \cdot X)^{-1}$). Соответствующий коэффициент уравнения регрессии считают зна-

чимым (т.е. значимо отличающимся от нуля), если $\left| \frac{b_i}{s(b_i)} \right| > t_\gamma$, где

t_γ — критическое значение распределения Стьюдента, определенное на уровне доверия $\gamma = 1 - q$ (где q — уровень значимости) при чис-

ле степеней свободы, равном $n - k - 1$ (т.е. квантиль уровня $\frac{\gamma}{2}$ распределения Стьюдента с $n - k - 1$ степенями свободы).

Более подробно с рассмотренными статистическими методами можно ознакомиться в литературных источниках [4—6].

Глава 2. МАТЕМАТИЧЕСКИЕ МЕТОДЫ В ТЕХНОЛОГИИ СТРОИТЕЛЬНЫХ МАТЕРИАЛОВ

2.1. Цель планирования эксперимента

Важной задачей методов обработки информации, полученной в ходе эксперимента, является задача построения математической модели изучаемого процесса, явления или объекта. Ее можно использовать и при проектировании объектов и при анализе процессов. Другой задачей обработки информации, полученной в ходе эксперимента, является задача оптимизации, т.е. нахождения такой комбинации влияющих независимых переменных, при которой выбранный показатель оптимальности принимает экстремальное значение.

Введем основные понятия, используемые в теории планирования эксперимента:

Опыт — это отдельная экспериментальная часть.

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

Планирование эксперимента — выбор плана эксперимента, удовлетворяющего заданным требованиям, или, иными словами, совокупность действий, направленных на разработку стратегии экспериментирования (от получения априорной информации до получения работоспособной математической модели или определения оптимальных условий). Это целенаправленное управление экспериментом, реализуемое в условиях неполного знания механизма изучаемого явления.

В процессе измерений, последующей обработки данных, а также формализации результатов в виде математической модели возникают погрешности и теряется часть информации, содержащей-

ся в исходных данных. Применение методов планирования эксперимента позволяет определить погрешность математической модели и судить о ее адекватности. Если точность модели оказывается недостаточной, то применение методов планирования эксперимента позволяет модернизировать математическую модель с проведением дополнительных опытов без потери предыдущей информации и с минимальными затратами.

Целью планирования эксперимента является нахождение таких условий проведения опытов, при которых удастся получить надежную и достоверную информацию об исследуемом объекте с наименьшей затратой труда, а также представить эту информацию с количественной оценкой точности.

2.2. Основные виды математических моделей

В основе обработки результатов активного и пассивного эксперимента с количественными факторами лежит *регрессионный анализ*. Он включает метод отыскания параметров математической модели (метод наименьших квадратов) и статистическую обработку данных.

Зависимость выходной величины (отклика) y от варьируемых факторов X_1, X_2, \dots, X_k , полученная с применением регрессионного анализа, называется *регрессионной моделью* и обозначается следующей формулой:

$$y = f(X_1, X_2, \dots, X_k), \quad (2.1)$$

где $f(X_1, X_2, \dots, X_k)$ — обозначение некоторой функции, называемой *функцией отклика*, от варьируемых факторов.

Регрессионная модель, таким образом, является частным случаем математической модели объекта. Выходных величин может быть несколько. Например, в процессе обработки детали на станке могут измеряться размеры детали, силы резания поверхности. Тогда зависимость вида (2.1) строится для каждого отклика.

Построенная регрессионная модель позволяет получить информацию о самом объекте и о способах управления им. С помощью регрессионной модели легко оценить степень и характер влияния каждого из факторов на выходную величину; модель может послужить основой для оптимизации процесса.

Вид регрессионной модели должен быть задан заранее, или до проведения эксперимента следует выбрать, к какому классу отно-

сится функция $f(X_1, X_2, \dots, X_k)$. Например, можно представить регрессионную модель в виде многочлена (полинома) определенного порядка, либо в виде экспоненты, тригонометрического многочлена и т.д.

Таким образом, при планировании эксперимента для математического описания объекта по результатам опытов рассчитывают только значения констант в регрессионной модели. Если, например, имеется единственный варьируемый фактор X_1 , а моделью является экспонента $Y = B_0 e^{B_1 X_1}$, то для построения модели в явном виде следует по результатам эксперимента вычислить значения коэффициентов B_0 и B_1 .

Возникает вопрос, как выбирается вид регрессионной модели. Здесь исследователю должны помочь знания об объекте, которыми он располагает до постановки эксперимента, — априорная информация (от латин. *apriori* — до опыта).

Поскольку вид регрессионной модели постулируется, задается до проведения эксперимента, пока остается открытым вопрос о достоверности такой модели. Для оценки применимости построенной модели, соответствия ее исследуемому объекту в планировании эксперимента предусмотрена специальная процедура, называемая *проверкой адекватности регрессионной модели*.

По результатам этой проверки исследователь имеет возможность принять или отвергнуть гипотезу о том, соответствует ли построенная модель результатам эксперимента, и, следовательно, пригодна ли она для описания объекта.

Наибольшее применение нашли методы планирования эксперимента, в которых регрессионные модели объектов представляются в виде многочленов (полиномов) первого и второго порядка от варьируемых факторов.

Модель в виде многочлена первого порядка сокращенно называют *регрессионной моделью первого порядка*, или *линейной моделью*. В общем случае, при наличии k варьируемых факторов линейная регрессионная модель объекта имеет вид

$$y = B_0 + B_1 X_1 + B_2 X_2 + \dots + B_k X_k, \quad (2.2)$$

где $B_0, B_1, B_2, \dots, B_k$ — коэффициенты, числовые значения которых определяются по результатам эксперимента. Их называют коэффициентами регрессии, а уравнение (2.2) или в общем случае уравнение (2.1) — уравнением регрессии. Коэффициенты B_1, B_2, \dots, B_k ,

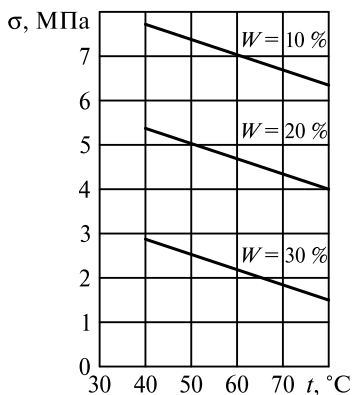


Рис. 2.1. График зависимостей $\sigma = f(t, W)$

стоящие перед обозначениями факторов X_1, X_2, \dots, X_k , называют линейными коэффициентами, а коэффициент B_0 — свободным членом.

Выбор регрессионной модели первого порядка для описания объекта равносильно предположению о линейной зависимости выходной величины от каждого из факторов, т.е. утверждению о том, что выходная величина изменяется пропорционально изменению варьируемого фактора.

Кроме того, представление регрессионной модели в виде многочлена первого порядка предполагает

отсутствие *эффектов взаимодействия* между факторами. Это означает, что степень и характер влияния каждого фактора на выходную величину не зависят от уровней варьирования остальных факторов.

На рис. 2.1 отсутствие эффектов взаимодействия между факторами t и W проявляется в параллельности прямых семейств.

Из всего этого следует, что линейная регрессионная модель дает, как правило, приближенное представление о влиянии факторов на объект. Применение таких моделей оправдано в следующих основных случаях:

- На начальных этапах исследования объекта или в других ситуациях, когда экспериментатора удовлетворяет ограниченная точность линейного приближения.
- При жестком ограничении количества опытов, поскольку экспериментальные планы, позволяющие получить линейную модель, являются экономными.
- В ситуации, когда экспериментатор уверен в достоверности линейной модели, например, по результатам теоретических исследований.

Модели второго порядка, т.е. модели в виде многочленов второго порядка от варьируемых факторов. Модель второго порядка (иначе — квадратичная модель) для трех варьируемых факторов имеет вид

$$y = B_0 + B_1X_1 + B_2X_2 + B_3X_3 + B_{11}X_1^2 + B_{22}X_2^2 + B_{33}X_3^2 + B_{12}X_1X_2 + B_{13}X_1X_3 + B_{23}X_2X_3. \quad (2.3)$$

Из уравнения (2.3) ясна общая структура квадратичной модели. Эта модель, рассматриваемая для произвольного числа k факторов, содержит все слагаемые линейной модели:

B_0 — свободный член;

$B_1X_1, B_2X_2, \dots, B_kX_k$ — линейные члены;

$B_{11}X_1^2, B_{22}X_2^2, \dots, B_{kk}X_k^2$ — квадратичные члены, являющиеся произведениями коэффициентов регрессии на квадраты факторов;

$B_{12}X_1X_2, B_{13}X_1X_3, \dots, B_{1k}X_1X_k, B_{23}X_2X_3, B_{24}X_2X_4, \dots, B_{2k}X_2X_k, \dots, B_{k-1k}X_{k-1}X_k$ — члены с парными взаимодействиями, которые представляют собой коэффициенты регрессии, умноженные на произведения двух различных факторов.

Зависимость выходной величины от каждого из факторов, полученная на основе квадратичной модели, представляется на графике отрезком параболы, имеющей ветви, направленные либо вверх (рис. 2.2, а), либо вниз (рис. 2.2, б). Такое представление позволяет достаточно полно описать широкий круг реальных зависимостей.

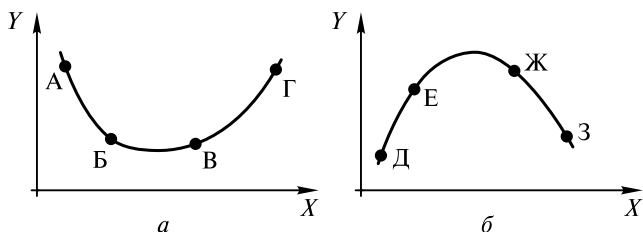


Рис. 2.2. Описание зависимостей моделями второго порядка

В случае монотонного возрастания выходной величины при росте варьируемого фактора (рис. 2.3, а, кривые 1 и 2) зависимость $y = f(X)$ будет представлена на графике участком ДЕ параболы (см. рис. 2.2, б) для кривой 1 или участком ВГ параболы (см. рис. 2.2, а) для кривой 2.

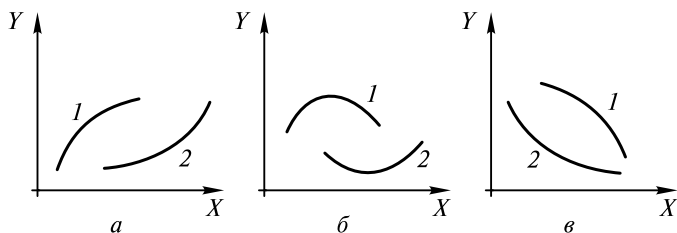


Рис. 2.3. Описание зависимостей моделями второго порядка

Монотонно убывающие зависимости (рис. 2.3, б, кривые 1 и 2) отобразятся при работе с квадратичной моделью участками парабол ЖЗ (см. рис. 2.2, б) для кривой 1 или АБ — для кривой 2.

Модель второго порядка удовлетворительно «работает» также при наличии одного максимума или минимума выходной величины в пределах диапазона варьирования факторов (рис. 2.3, в). Этому случаю соответствуют участки парабол ЕЖ и БВ на рис. 2.2.

Описание объекта квадратичной моделью даст заведомо плохие результаты, если:

- истинная зависимость отклика от некоторого фактора X_i имеет более одного экстремума;
- зависимость $y = f(X)$ имеет точку перегиба;
- при некотором значении X_i значение отклика резко (скачком) изменяется.

В первых двух случаях можно рекомендовать для описания объекта многочлены (полиномы) третьего или более высокого порядка. Другим выходом из положения, пригодным для всех трех случаев, является деление диапазона варьирования факторов на более мелкие поддиапазоны и изучение объекта для каждой из полученных областей отдельно.

2.3. Метод наименьших квадратов

2.3.1. МЕТОД НАИМЕНЬШИХ КВАДРАТОВ ДЛЯ МОДЕЛЕЙ С ОДНОЙ ПЕРЕМЕННОЙ

Рассмотрим идею основного (для количественных факторов) метода обработки результатов эксперимента с целью получения математического описания объекта — *метода наименьших квадратов*.

Рассмотрим случай варьирования единственного фактора X_j . Предположим, что эксперимент состоит в постановке N опытов, и в этих опытах фактор X_1 принимает значения $X_{11}, X_{12}, \dots, X_{1N}$. Здесь X_{1j} — значение фактора X_1 в опыте номер $j, j = 1, 2, \dots, N$. Выходная величина y принимает в этих опытах значения y_1, y_2, \dots, y_N , соответственно.

Отложим по оси абсцисс значения фактора X_1 , принимаемые им в опытах, а по оси ординат — соответствующие значения y , получим совокупность точек (рис. 2.4). Цель эксперимента — получение регрессионной модели $y = f(X_1)$, которая с достаточной точностью описывала бы результаты эксперимента.

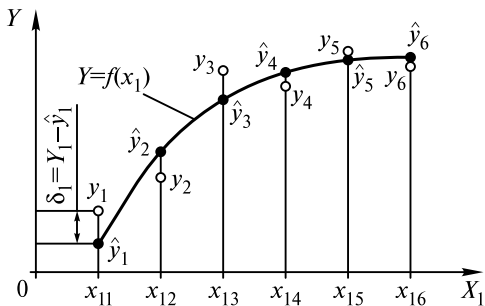


Рис. 2.4. Регрессионная модель $y = f(X_1)$

Требуется исследовать зависимость прочности бетона от доли песка в смеси. Доля песка оценивается отношением $r = \Pi / (\Pi + \text{Щ})$. Тогда точки y_1, y_2, \dots, y_6 на рис. 2.4 это значение r , соответствующее разным значениям продолжительности резания. Закономерность изменения прочности в зависимости от доли песка получим на графике, если проведем гладкую кривую, лежащую возможно ближе к экспериментальным точкам. Однако ориентировочно такую кривую можно провести разными способами и, кроме того, помимо графика хотелось бы получить аналитическое представление для исследуемой зависимости. Все это заставляет обратиться к аналитическим методам построения регрессионной модели.

Конкретизируем приведенное выше требование, чтобы экспериментальные точки лежали в совокупности как можно ближе к кривой, являющейся графиком искомой зависимости.

Допустим, что аналитическое представление зависимости y от X_1 уже каким-то образом получено в виде уравнения регрессии $y = f(X_1)$. График зависимости $y = f(X_1)$ это искомая кривая (см. рис. 2.4).

Значениям фактора X_1 , равным $X_{11}, X_{12}, \dots, X_{1N}$, соответствуют точки на кривой $\hat{y}_1, \hat{y}_2, \dots, \hat{y}_N$. Эти точки являются значениями выходной величины, рассчитанными по уравнению регрессии $y = f(X_1)$:

$$\hat{y}_1 = f(X_{11});$$

$$\hat{y}_2 = f(X_{12});$$

.....

$$\hat{y}_N = f(X_{1N}).$$

Найдем величину δ_1 , равную $\delta_1 = y_1 - \hat{y}_1$ (см. рис. 2.4), которая характеризует отклонение результата эксперимента y_1 в точке X_{11} от значения функции отклика $\hat{y}_1 = f(X_{11})$ в этой же точке. Аналогично рассмотрим отклонения $\delta_1 = y_1 - \hat{y}_1, \dots, \delta_N = y_N - \hat{y}_N$.

Согласно методу наименьших квадратов (МНК), оценки для коэффициентов регрессии находятся из условия минимума суммы квадратов отклонений Φ , т.е.

$$\Phi = \delta_1^2 + \delta_2^2 + \dots + \delta_N^2 = (y_1 - \hat{y}_1)^2 + (y_2 - \hat{y}_2)^2 + \dots + (y_N - \hat{y}_N)^2 = \sum_{j=1}^n (y_j - \hat{y}_j)^2 \rightarrow \min.$$

Рассмотрим вычисление *коэффициентов квадратичной модели* с единственным фактором X_1 :

$$y = B_0 + B_1 X_1 + B_{11} X_1^2.$$

Для отыскания трех неизвестных коэффициентов регрессии B_0, B_1 и B_{11} надо решить следующую систему из трех линейных уравнений с тремя неизвестными:

$$\begin{cases} NB_0 + B_1 \sum X_{1j} + B_{11} \sum X_{1j}^2 = \sum y_j, \\ B_0 \sum X_{1j} + B_1 \sum X_{1j}^2 + B_{11} \sum X_{1j}^3 = \sum y_j X_{1j}, \\ B_0 \sum X_{1j}^2 + B_1 \sum X_{1j}^3 + B_{11} \sum X_{1j}^4 = \sum y_j X_{1j}^2. \end{cases}$$

При равноотстоящих значениях фактора X_1 математическую модель удобно представить в виде

$$y = B'_{11} \left(\frac{X_1 - \bar{X}_1}{h} \right)^2 + B'_1 \left(\frac{X_1 - \bar{X}_1}{h} \right) + B'_0, \quad (2.4)$$

где h — шаг варьирования фактора X_1 ;

B'_{11}, B'_1, B'_0 — коэффициенты, вычисляемые по следующим формулам:

При нечетном N :

$$B'_{11} = \frac{1}{3H_2} \left\{ \begin{array}{l} 3[y_1(1-M)^2 + y_2(2-M)^2 + \dots + y_N(N-M)^2] - \\ - \frac{N^2-1}{4}(y_1 + y_2 + \dots + y_N) \end{array} \right\}; \quad (2.5)$$

$$B'_1 = \frac{1}{H_1} [y_1(1-M) + y_2(2-M) + \dots + y_N(N-M)]; \quad (2.6)$$

$$B'_0 = \bar{y} - (1H_1)B'_{11}, \quad (2.7)$$

где $H_2 = N - (N^2 - 1) \frac{(N^2 - 4)}{180}$;

$$H_1 = N - \frac{(N^2 - 1)}{12};$$

$$M = (N + 1) / 2.$$

При четном N :

$$B'_{11} = \frac{1}{12H_2} \left\{ \begin{array}{l} 3 \left[\begin{array}{l} y_1(2-N-1)^2 + y_2(4-N-1)^2 + \\ + y_3(6-N-1)^2 + \dots + y_N(2N-N-1)^2 \end{array} \right] - \\ - (N^2-1)(y_1 + y_2 + \dots + y_N) \end{array} \right\};$$

$$B'_1 = 2 \frac{1}{H_1} [y_1(2-N-1) + y_2(4-N-1) + \dots + y_N(2N-N-1)];$$

$$B'_0 = \bar{y} - (H_1/N)B'_{11}, \quad (2.8)$$

где $H_2 = N - (N^2 - 1) \frac{(N^2 - 4)}{180}$;

$$H_1 = N - \frac{(N^2 - 1)}{12};$$

$$M = N / 2.$$

Пример 2.1. В эксперименте получены данные о пределе прочности бетона в возрасте 28 суток в зависимости от доли песка в смеси заполнителей r (табл. 2.1).

Требуется рассчитать значение прочности бетона в возрасте 28 суток в зависимости от доли песка в смеси заполнителей и построить график зависимости предела прочности бетона в этом возрасте от доли песка в смеси заполнителей.

Таблица 2.1

К примеру 2.1

$r = \Pi/(\Pi+\Щ)$	R_{28} , МПа	$r = \Pi/(\Pi+\Щ)$	R_{28} , МПа
0,25	44,7	0,40	30,2
0,3	36,3	0,45	39,1
0,35	42,5	—	—

Примечание. П — песок; Щ — щебень.

Решение

Опишем эту зависимость уравнением параболы в виде уравнения (2.4). Фактор $r \equiv X_1$ варьируется с постоянным шагом $h = 0,05$. Поскольку число опытов $N = 5$ нечетно, коэффициенты B'_{11} , B'_1 , B'_0 найдем по формулам (2.5)—(2.7), соответственно.

Предварительно рассчитываем:

$$\bar{X} = 0,35;$$

$$\bar{y} = \frac{192,8}{5} = 38,56;$$

$$M = \frac{N+1}{2} = 3;$$

$$H_1 = N - \frac{(N^2-1)}{12} = 10;$$

$$H_2 = N - (N^2-1) \frac{(N^2-4)}{180} = 14;$$

$$B'_{11} = \frac{1}{3 \cdot 14} \left\{ \begin{array}{l} 3 \left[44,7(1-3)^2 + 36,3(2-3)^2 + 42,5(3-3)^2 + \right. \\ \left. + 30,2(4-3)^2 + 39,1(5-3)^2 \right] - \\ \left. - \frac{5^2-1}{4} (44,7+36,3+42,5+30,2+39,1) \right\} = 1,15;$$

$$B'_1 = \frac{1}{10} \left\{ \left[44,7(1-3) + 36,3(2-3) + 42,5(3-3) + \right. \right. \\ \left. \left. + 30,2(4-3) + 39,1(5-3) \right] \right\} = -1,73;$$

$$B'_0 = 38,56 - (1/10)B'_1 = 38,45.$$

В результате получим уравнение

$$y = 1,15 \left(\frac{X_1 - 0,35}{0,05} \right)^2 - 1,73 \left(\frac{X_1 - 0,35}{0,05} \right) + 38,45.$$

Возвращаясь к исходным обозначениям, после преобразований будем иметь:

$$y = -11,6X_1^2 + 8,8X_1 + 37,0.$$

Рассчитанные по уравнению регрессии значения прочности приведены в табл. 2.2, на основании этих значений построен график зависимости предела прочности бетона в возрасте 28 суток от доли песка в смеси заполнителей (рис. 2.5).

Таблица 2.2

К примеру 2.1

$r = \Pi / (\Pi + \text{Щ})$	R_{28} , МПа	$r = \Pi / (\Pi + \text{Щ})$	R_{28} , МПа
0,25	38,31	0,40	38,58
0,3	38,36	0,45	38,31
0,35	38,41	—	—

Примечание. П — песок; Щ — щебень.

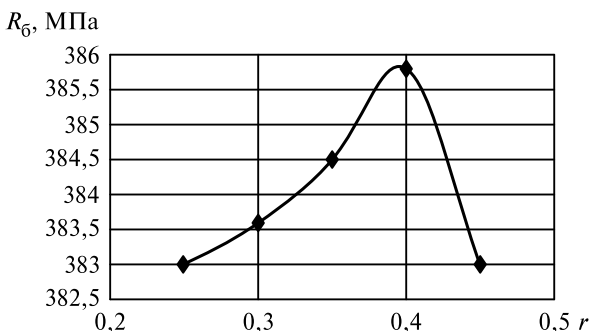


Рис. 2.5. График зависимости предела прочности бетона в возрасте 28 суток от доли песка в смеси заполнителей

2.3.2. МЕТОД НАИМЕНЬШИХ КВАДРАТОВ ДЛЯ МНОГОФАКТОРНЫХ ЭКСПЕРИМЕНТОВ

Линейная регрессионная модель с k варьируемыми факторами имеет следующий вид:

$$y = B_0 + B_1X_1 + B_2X_2 + \dots + B_kX_k.$$

Значения факторов, принимаемые в каждом опыте, можно свести в табл. 2.3.

Таблица 2.3

Матрица планирования эксперимента

№ опыта	Значения фактора			
	X_1	X_2	...	X_k
1	X_{11}	X_{21}	...	X_{k1}
2	X_{12}	X_{22}	...	X_{k2}
3	X_{13}	X_{23}	...	X_{k3}
.
N	X_{1N}	X_{2N}	...	X_{kN}

Таблицы, составленные по типу табл. 2.3, в которых записаны условия опытов, называются *матрицами* планов. Прежде, чем проводить вычисления, запишем регрессионную модель в более симметричном виде, введя фиктивный фактор X_0 :

$$y = B_0X_0 + B_1X_1 + B_2X_2 + \dots + B_kX_k. \quad (2.9)$$

В связи с этим дополним матрицу плана в табл. 2.3, введя в нее столбец значений фиктивного фактора X_0 (табл. 2.4).

Таблица 2.4

Матрица планирования эксперимента

№ опыта	Значения фактора				
	X_0	X_1	X_2	...	X_k
1	X_{01}	X_{11}	X_{21}	...	X_{k1}
2	X_{02}	X_{12}	X_{22}	...	X_{k2}
3	X_{03}	X_{13}	X_{23}	...	X_{k3}
.
N	X_{0N}	X_{1N}	X_{2N}	...	X_{kN}

Теперь каждому слагаемому модели (2.9) соответствует определенный столбец этой матрицы: слагаемому B_0X_0 — столбец X_0 , слагаемому B_1X_1 — столбец X_1 , и т.д.

Такая матрица называется *матрицей базисных функций*. В табл. 2.4 таким образом построена матрица базисных функций модели (2.9) для плана табл. 2.3.

Для нахождения коэффициентов регрессии B_0, B_1, \dots, B_k модели используем систему нормальных уравнений:

$$\begin{cases} B_0 \sum X_{0j}^2 + B_1 \sum X_{0j}X_{1j} + B_2 \sum X_{0j}X_{2j} + \dots + B_k \sum X_{0j}X_{kj} = \sum X_{0j}y_j, \\ B_0 \sum X_{0j}X_{1j} + B_1 \sum X_{1j}^2 + B_2 \sum X_{1j}X_{2j} + \dots + B_k \sum X_{1j}X_{kj} = \sum X_{1j}y_j, \\ B_0 \sum X_{0j}X_{2j} + B_1 \sum X_{2j}X_{1j} + B_2 \sum X_{2j}^2 + \dots + B_k \sum X_{12j}X_{kj}, \\ \dots \\ B_0 \sum X_{0j}X_{kj} + B_1 \sum X_{kj}X_{1j} + B_2 \sum X_{kj}X_{2j} + \dots + B_k \sum X_{kj}^2 = \sum X_{kj}y_j. \end{cases} \quad (2.10)$$

Система уравнений (2.10) построена по тому же принципу, что и система нормальных уравнений для линейной однофакторной модели. Ее легко написать, имея матрицу базисных функций.

Отметим, что число уравнений этой системы равно числу коэффициентов регрессии, подлежащих определению, т.е. в данном случае $k + 1$.

Ясно также, что в случае применения метода наименьших квадратов число опытов N должно быть не меньше числа p оцениваемых коэффициентов регрессии: $N \geq p$.

План, для которого $p = N$, называется *насыщенным* планом. Насыщенные планы не позволяют проверить адекватность математической модели.

План, для которого $p < N$, называется *ненасыщенным*.

Расчет коэффициентов регрессии и интерпретация результатов эксперимента существенно упрощаются, если преобразовать все факторы в безразмерные параметры, варьируемые в одинаковых диапазонах. Это можно сделать, введя нормализованные обозначения факторов.

Пример 2.2. По данным эксперимента водоцементное отношение бетонной смеси изменяется в пределах 0,4—0,6, доля песка в смеси заполнителей — в пределах 0,35—0,45.

Требуется получить зависимость предела прочности бетона в возрасте 28 суток нормального твердения от В/Ц (X_1) и доли песка в смеси заполнителей r (X_2).

Решение

Выберем интервалы варьирования факторов.

Согласно экспериментальному плану, фактор X_1 и фактор X_2 варьируются на нижнем, верхнем и основном уровне (табл. 2.5, 2.6).

Таблица 2.5

Факторы и уровни варьирования эксперимента

Факторы		Уровни варьирования			Интервал варьирования
Нормализованное (кодовое) значение	Натуральное значение	Верхний (+1)	Средний (0)	Нижний (-1)	
X_1	В/Ц	0,6	0,5	0,4	0,1
X_2	$r = П/(П+Щ)$	0,45	0,40	0,35	0,05

Таблица 2.6

План эксперимента и матрица в нормализованных значениях

№ опыта	Нормализованные значения			Натуральные значения			Выходные параметры	
	x_0	x_1	x_2	X_0	$X_1 \equiv В/Ц$	$X_2 \equiv r$	$y \equiv R_{28}, \text{ МПа}$	\hat{y}
1	+1	-1	-1	1	0,4	0,35	42,5	46,8
2	+1	0	-1	1	0,5	0,35	35,3	42,7
3	+1	+1	+1	1	0,6	0,45	30,2	38,4
4	+1	-1	+1	1	0,4	0,45	39,1	44,6
5	+1	0	0	1	0,5	0,40	37,5	41,6
6	+1	+1	0	1	0,6	0,40	34,0	39,5
	6	0	0	6	3	2,4	218,6	

Для составления системы нормальных уравнений вычислим суммы произведений элементов каждой пары столбцов — см. систему уравнений (2.15) (табл. 2.7).

Таблица 2.7

К примеру 2.2

$\sum x_{0j}^2$	$\sum x_{0j}x_{1j}$	$\sum x_{0j}x_{2j}$	$\sum x_{0j}y_j$	$\sum x_{1j}^2$	$\sum x_{1j}x_{2j}$	$\sum x_{1j}y_j$	$\sum x_{2j}^2$	$\sum x_{2j}y_j$
6	0	0	218,6	4	1	-17,4	4	-8,5

Рассчитаем коэффициент регрессии по системе уравнений (2.10):

$$\begin{cases} 6b_0 = 218,6 & \rightarrow b_0 = 36,43 \\ 4b_1 + b_2 = -17,4 & \rightarrow b_1 = -4,07 \\ b_1 + 4b_2 = -8,5 & \rightarrow b_2 = -1,12 \end{cases}$$

Таким образом, получена следующая линейная модель для нормализованных значений:

$$y = 36,43 + 4,07x_1 - 1,12x_2.$$

Для иллюстрации точности, с которой построенная модель предсказывает результаты эксперимента, в последнем столбце табл. 2.4 приведены значения отклика \hat{y} , рассчитанные по уравнению регрессии для каждого опыта. При переходе к натуральным факторам следует воспользоваться полученными выше формулами, связывающими нормализованные и натуральные обозначения факторов.

Подставив в найденное уравнение регрессии значения

$$x_i = (X_i - X_i^{(0)})/\Delta_i;$$

$$x_1 = (X_1 - 0,6)/0,1;$$

$$x_2 = (X_2 - 0,45)/0,05,$$

получим

$$y = 36,43 + 4,07(X_1 - 0,6)/0,1 - 1,12(X_2 - 0,45)/0,05.$$

После преобразования будем иметь математическую модель

$$y = 60,85 - 40,7X_1 - 22,4X_2.$$

2.3.3. СТАТИСТИЧЕСКИЙ АНАЛИЗ УРАВНЕНИЯ РЕГРЕССИИ

Дисперсия воспроизводимости

После того как получено уравнение регрессии, проводят его статистический анализ. При этом решают две задачи:

- оценивают значимость коэффициентов регрессии;
- проверяют адекватность математической модели.

Для выполнения каждой из этих задач необходимо иметь количественную оценку ошибок эксперимента в целом. Соответствующей характеристикой является дисперсия воспроизводимости, которая обозначается $s^2\{y\}$. Рассмотрим способы ее вычисления в зависимости от методики дублирования опытов.

1. *Равномерное дублирование.* Каждый из N запланированных опытов повторяется одинаковое число n раз, т.е. имеется N серий, в каждой из которых ставится n дублированных опытов.

По результатам опытов первой серии рассчитываем дисперсию первого опыта:

$$s_1^2 = \frac{[(y_{11} - \bar{y}_1)^2 + (y_{12} - \bar{y}_1)^2 + \dots + (y_{1n} - \bar{y}_1)^2]}{(n-1)} = \sum \frac{(y_{1u} - \bar{y}_1)^2}{(n-1)},$$

где \bar{y}_1 — среднее по серии дублированных опытов, равное

$$\bar{y}_1 = \frac{y_{11} + y_{12} + \dots + y_{1n}}{n} = \sum \frac{y_{1u}}{n}.$$

Аналогично рассчитываются средние и дисперсии всех остальных опытов:

$$\bar{y}_j = \sum y_{ju}/n \text{ и } s_j^2 = \sum (y_{ju} - \bar{y}_j)^2 / (n-1) \text{ при } j = 1, 2, \dots, N.$$

Числа степеней свободы всех дисперсий одинаковы и равны $n - 1$ $f_j = f = n - 1$.

В качестве дисперсии воспроизводимости $s^2\{y\}$ берется среднее арифметическое дисперсий опытов:

$$s^2\{y\} = \frac{s_1^2 + s_2^2 + \dots + s_N^2}{N} = \sum s_j^2 / N.$$

Число степеней свободы f_y этой дисперсии равно сумме чисел степеней свободы дисперсий опытов:

$$f_y = \sum f_j = N(n - 1).$$

Необходимыми предпосылками статистического анализа являются нормальность распределения выходной величины и однородность дисперсий опытов.

Проверка однородности дисперсий опытов при равномерном дублировании проводится по критерию Кохрена.

2. *Отсутствие дублированных опытов (опыты в нулевых точках).* Для оценки дисперсии воспроизводимости в этом случае приходится ставить отдельную серию дублированных опытов, если это возможно.

Как и в предыдущем случае, дисперсия опытов этой серии служит оценкой дисперсии воспроизводимости с числом степеней свободы, равным $f_y = n_0 - 1$, где n_0 — число дублированных опытов в отдельной серии.

Оценка точности и значимости коэффициентов регрессии и интерпретация результатов этой оценки

Статистическую обработку проводят обычно для модели, записанной в нормализованных обозначениях факторов. Для определенности будем иметь в виду линейную модель, содержащую k факторов.

После того, как уравнение регрессии получено и рассчитана дисперсия воспроизводимости, нужно оценить точность, с которой найдены коэффициенты регрессии.

Так как коэффициенты вычислены по результатам эксперимента, а эти результаты являются случайными величинами, то и коэффициенты регрессии b_i будут случайными величинами.

Поэтому в качестве показателя точности отыскания коэффициента b_i удобно взять его дисперсию $s\{b_i\}$.

Для большинства планов, рекомендованных теорией эксперимента, существуют простые формулы для отыскания дисперсий коэффициентов регрессии (будут рассмотрены дальше при изучении планов эксперимента).

После того как найдены дисперсии коэффициентов регрессии, необходимо выявить незначимые коэффициенты, т.е. те, которые в математической модели можно приравнять нулю.

Для этого используется t -критерий Стьюдента. Для каждого коэффициента регрессии b_i отыскивается t -отношение

$$t_i = |b_i|/s\{b_i\}. \quad (2.11)$$

Вычисленную величину t_i сравнивают с табличным значением $t_{\text{табл}}$ t -критерия Стьюдента для заданного уровня значимости q и

числа степеней свободы f_y , с которым определялась дисперсия воспроизводимости $s\{y\}$.

Если $t_i < t_{\text{табл}}$, то коэффициент регрессии b_i незначим и соответствующий член в уравнении регрессии должен быть отброшен.

С учетом (2.11) условие того, что коэффициент регрессии незначим, можно записать в более удобном виде

$$|b_i| \leq t_{\text{табл}} s\{b_i\}. \quad (2.12)$$

При отбрасывании незначимых членов возникает определенное неудобство, связанное со статической зависимостью коэффициентов регрессии. Эта зависимость проявляется в том, что, после того как незначимые коэффициенты регрессии приравняли нулю, оценки остальных коэффициентов регрессии изменяются.

Практический вывод: после отбрасывания незначимых коэффициентов регрессии желательно снова воспользоваться МНК для уточнения оставшихся значимых коэффициентов регрессии.

С помощью t -критерия можно найти и доверительный интервал для произвольного коэффициента регрессии b_i . Обозначим истинную величину этого коэффициента через β_i . Тогда

$$b_i - t_{\text{табл}} s\{b_i\} \leq \beta_i \leq b_i + t_{\text{табл}} s\{b_i\},$$

и даже простейшая линейная модель позволяет получить важную информацию об объекте исследования.

Запишем линейную модель в нормализованных обозначениях факторов:

$$y = b_0 + b_1 x_1 + b_2 x_2 + \dots + b_k x_k. \quad (2.13)$$

Коэффициенты этой модели имеют четкий физический смысл:

- Коэффициент b_0 равен, очевидно, значению выходной величины, рассчитанному по уравнению регрессии, если все факторы зафиксированы на основном уровне, т.е. в середине диапазона варьирования.

- Знак коэффициента b_i свидетельствует о характере влияния соответствующего фактора:

- если $b_i > 0$, то с ростом значений факторов X_i выходная величина растет;

- если $b_i < 0$, то с ростом X_i отклик уменьшается.

- Величина b_i равна приросту выходной величины, полученному при увеличении значения фактора X_i на половину диапазона его варьирования, например, с основного уровня ($X_i = X_i^0$) до верхнего уровня ($X_i = X_i^1$).

- Графиком зависимости величины y от любого фактора будет прямая.

- Рассмотрение зависимостей выходной величины от любого фактора при разных фиксированных значениях других факторов позволит получить семейство прямых, причем все эти прямые будут параллельны. Это связано с тем, что представление регрессионной модели в линейном виде предполагает отсутствие взаимодействий факторов.

- Чем больше абсолютная величина линейного коэффициента регрессии в модели (2.13), тем сильнее влияние соответствующего фактора. Если, например, оказалось, что $|b_3| > |b_1|$, то можно сделать вывод о том, что изменение фактора X_3 в пределах его диапазона варьирования оказывает большее влияние на изменение отклика, чем варьирование фактора X_1 в его диапазоне.

С помощью линейной регрессионной модели можно сравнить степень влияния факторов на выходную величину и выявить важнейшие факторы.

Если уравнение регрессии отличается от линейного, то степень влияния фактора может изменяться от начала к концу диапазона варьирования и зависит от уровней варьирования других факторов.

Проверка адекватности регрессионной модели

Регрессионная модель, построенная по результатам эксперимента, позволяет рассчитать значения отклика в разных точках области варьирования факторов.

Для этого в уравнение регрессии подставляют соответствующие значения варьируемых факторов.

Проверка адекватности математической модели дает возможность экспериментатору ответить на вопрос, будет ли построенная модель предсказывать значения выходной величины с той же точностью, что и результаты эксперимента.

Пусть N — число опытов экспериментального плана или число серий параллельных опытов, если опыты дублируются; p — число оцениваемых коэффициентов регрессии математической модели.

Проверка адекватности возможна только при $N > p$, т.е. если план эксперимента является насыщенным.

Для проверки адекватности модели необходимо знать оценку дисперсии воспроизводимости $s^2\{y\}$, которую можно вычислить в зависимости от методики дублирования по одной из формул, приведенных ранее в подразд. 2.3.3.

Порядок проверки адекватности модели следующий:

- Определяют сумму квадратов, характеризующую адекватность модели $s_{ад}$. При равномерном дублировании ее рассчитывают по формуле

$$s_{ад} = n[(\bar{y}_1 - \hat{y}_1)^2 + \bar{y}_2 - \hat{y}_2)^2 + \dots + \bar{y}_N - \hat{y}_N)^2] = n \sum (\bar{y}_j - \hat{y}_j)^2,$$

где n — число дублированных опытов в каждой серии; \bar{y}_j — среднее значение результатов эксперимента в j -й серии дублированных опытов, $j = 1, 2, \dots, N$; \hat{y}_j — среднее значение по уравнению регрессии для j -го основного опыта.

В случае неравномерного дублирования

$$s_{ад} = \sum n_j (\bar{y}_j - \hat{y}_j)^2,$$

где n — число дублированных опытов в j -й серии.

- Вычисляют число степеней свободы $f_{ад}$ дисперсии адекватности. При любой методике дублирования опытов оно равно

$$f_{ад} = N - p.$$

- Вычисляют дисперсию адекватности $s_{ад}^2$:

$$s_{ад}^2 = s_{ад} / f_{ад}. \quad (2.14)$$

- С помощью F -критерия Фишера проверяют однородность дисперсии адекватности $s_{ад}^2$ и дисперсии воспроизводимости $s^2\{y\}$.

При этом вычисляют

$$F_{расч} = s_{ад}^2 / s^2\{y\},$$

которое сравнивают с табличным значением F -критерия $F_{табл}$, найденным при выбранном уровне значимости q для чисел степеней свободы $f_{ад}$ в числителе и f_y в знаменателе (см. табл. П4 приложения).

Если $F_{\text{расч}} < F_{\text{табл}}$, то модель считается адекватной и может быть использована для описания объекта. В противном случае модель неадекватна.

Если оказалось, что уравнение неадекватно, то при проведении опытов были допущены грубые ошибки или выбранный полином недостаточно полно отражает исследуемую зависимость. В этих случаях необходимо повторить опыты либо изменить интервалы варьирования, либо применить другой план.

Рассмотренный метод проверки адекватности модели имеет простой физический смысл. В основе этой процедуры лежит проверка гипотезы об однородности дисперсии адекватности и дисперсии, характеризующей ошибку эксперимента.

Заметим, что дисперсия адекватности характеризует расхождение между результатами эксперимента y_j и значениями выходной величины \hat{y}_j , вычисленными по уравнению регрессии.

Логично принять, что модель удовлетворительно описывает объект исследования, т.е. является адекватной, если указанное расхождение связано только экспериментальными ошибками, а не связано, например, с неудачным выбором вида математической модели.

Проверка гипотезы об однородности рассматриваемых дисперсий и выясняет «общность происхождения» экспериментальных ошибок и расхождения между y_j и \hat{y}_j .

Кроме проверки адекватности модели можно оценить ее эффективность, информационную ценность.

При отсутствии дублированных опытов эффективность регрессионной модели оценивают следующим образом:

- Вычисляют дисперсию относительно среднего значения отклика:

$$s_c^2 = \sum (y_j - \bar{y})^2 / (N - 1),$$

где \bar{y} — среднее значение отклика по всем опытам; $\bar{y} = \sum y_i / N$.

- Рассчитывают остаточную дисперсию $s_{\text{ост}}^2$:

$$s_{\text{ост}}^2 = \sum (y_j - \hat{y}_j)^2 / (N - p).$$

- Вычисляют отношение $F_{\text{расч}}$

$$F_{\text{расч}} = s_c^2 / s_{\text{ост}}^2. \quad (2.15)$$

Величина $F_{\text{расч}}$ показывает, во сколько раз уравнение регрессии описывает результаты эксперимента точнее, чем простое среднее арифметическое, взятое по всем опытам.

Регрессионная модель считается эффективной, если $F_{\text{расч}} > (3-5)$.

Для экспериментов с дублированными опытами формула (2.15) остается в силе, а выражения для дисперсий s_c^2 и $s_{\text{ост}}^2$ примут следующий вид:

$$s_c^2 = \sum_{j=1}^N \sum_{u=1}^{n_j} (y_{ju} - \bar{y})^2 / \left(\sum_{j=1}^N n_j - 1 \right);$$

$$s_{\text{ост}}^2 = \sum_{j=1}^N \sum_{u=1}^{n_j} (y_{ju} - \hat{y})^2 / \left(\sum_{j=1}^N n_j - p \right),$$

где y_{ju} — значение отклика в u -м дублированном опыте j -й серии; n — число серий дублированных опытов;

$$\bar{y} = \sum_{j=1}^N \sum_{u=1}^{n_j} y_{ju} / \sum_{j=1}^N n_j.$$

Последовательность действий при проведении эксперимента с целью построения регрессионной модели объекта следующая:

- Выбор варьируемых и стабилизируемых факторов, а также выходных величин эксперимента.
- Выбор регрессионной модели.
- Определение диапазона варьирования факторов.
- Выбор плана эксперимента.
- Составление методики проведения эксперимента.
- Постановка разведывательных опытов. Проверка нормальности распределения выходной величины. Определение числа дублированных опытов.
- Проведение основного эксперимента.
- Отбрасывание грубых наблюдений. Проверка однородности дисперсий опытов. Расчет дисперсии воспроизводимости (при отсутствии дублированных опытов дисперсия воспроизводимости определяется по результатам отдельной серии опытов).
- Расчет коэффициентов регрессии математической модели.

- Оценка значимости коэффициентов регрессии. Проверка адекватности и эффективности регрессионной модели.
- Интерпретация результатов.

Приведенный перечень этапов только приблизительно отражает реальную последовательность действий экспериментатора, поскольку многие этапы оказываются взаимосвязанными. Таковы, например, выбор математической модели и определение диапазона варьирования факторов.

При выборе диапазона варьирования факторов существенны, прежде всего, соображения экспериментатора, связанные с возможностью применения полученных им результатов и рекомендаций в производственной сфере. Поэтому диапазоны варьирования факторов в эксперименте обычно соответствуют реальным производственным условиям.

Необходимо отметить, кроме того, что диапазоны варьирования факторов следует выбирать тем больше, чем ниже точность фиксирования факторов и чем меньше диапазон изменения выходной величины. С другой стороны, для адекватного описания объекта моделью второго порядка требуются, по сравнению с моделью первого порядка, более узкие диапазоны варьирования факторов.

Глава 3. ЭКСПЕРИМЕНТАЛЬНЫЕ ПЛАНЫ ПЕРВОГО ПОРЯДКА

3.1. Этапы планирования эксперимента

Планирование эксперимента начинается со сбора исходных данных, позволяющих получить максимальную информацию о параметре оптимизации. Информация для этого берется как из литературных источников, так и из результатов сторонних исследований по подобным направлениям. Основной целью сбора информации является определение экспериментальной области факторного пространства и выбор основного (нулевого) уровня и интервалов варьирования факторов. В большинстве случаев собранной информации недостаточно для определения нулевого уровня и интервалов факторов варьирования. Поэтому на данном этапе необходимо проведение однофакторных экспериментов, позволяющих определить граничные (допустимые) значения выбранных факторов.

Исходной точкой для построения плана эксперимента является основной (нулевой) уровень, который определяется как середина отрезка, построенного между точками, указывающими предельные значения варьирования фактора.

При планировании эксперимента значения факторов кодируются. Кодирование факторов производится путем линейного преобразования координат факторного пространства, выбора масштабов по осям в единицах интервалов варьирования факторов и переноса начала координат в нулевую точку.

Значения факторов после данного преобразования равны $+1$ (верхний уровень) и -1 (нижний уровень).

Положение точек проведения эксперимента в факторном пространстве для полного факторного эксперимента при $k = 2$ и $k = 3$ показано на рис. 3.1.

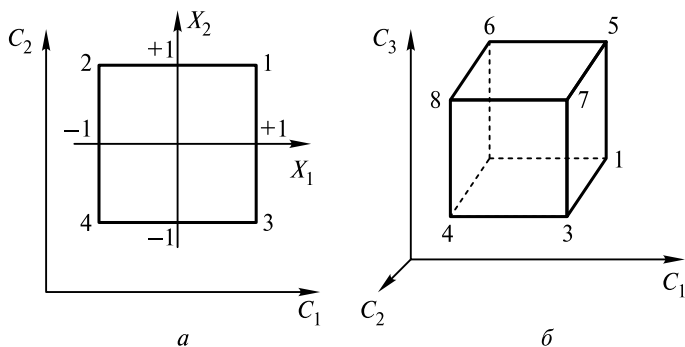


Рис. 3.1. Расположение точек в факторном пространстве для ПФП при $k = 2$ (а) и $k = 3$ (б)

Как показано на рис. 3.1, а, точки плана 2^2 задаются координатами вершин квадрата, а точки плана 2^3 (рис. 3.1, б) — координатами вершин куба. Данный принцип применяется при расположении экспериментальных точек при $k > 3$.

3.2. Полный факторный план типа 2^k

Начальный этап планирования эксперимента, в результате которого будет получена линейная модель, осуществляется при варьировании факторов на двух уровнях. В данном случае, при определенном количестве факторов, находится число опытов, необхо-

димых для реализации всех возможных сочетаний уровней факторов.

Расчет числа опытов проводится по формуле $N = p^k$ (где N — число опытов; p — число уровней; k — число факторов) и в рассматриваемом случае определяется как $N = 2^2$.

Полным факторным планом (ПФП) называется эксперимент, в котором реализуются все возможные сочетания уровней факторов. ПФП с двумя уровнями факторов имеет вид 2^k .

Отдельные эксперименты во всех возможных сочетаниях наиболее практично обозначать в виде таблицы, которая называется матрицей планирования эксперимента. Матрица планирования эксперимента для двух факторов приведена в табл. 3.1.

Таблица 3.1

Матрица планирования эксперимента 2^2

№ опыта	x_1	x_2	y
1	-1	-1	y_1
2	-1	+1	y_1
3	+1	-1	y_3
4	+1	+1	y_4

При заполнении матрицы планирования значения уровней факторов в целях упрощения в матрицах планирования полных факторных экспериментов цифру 1 опускают, а уровни обозначают знаками «+» и «-». С учетом указанного упрощения табл. 3.1 можно переписать в виде табл. 3.2.

Таблица 3.2

Матрица планирования эксперимента 2^2

№ опыта	x_1	x_2	y
1	-	-	y_1
2	-	+	y_1
3	+	-	y_3
4	+	+	y_4

Столбцы в матрице планирования называют векторами-столбцами, а строки — векторами-строками. В табл. 3.1 и 3.2 присутствуют два вектора-столбца независимых переменных и один вектор-столбец параметра оптимизации.

При построении точек в области определения факторов необходимо найти точку, соответствующую основному уровню. Через данные точки проводят новые оси координат, параллельные осям натуральных значений факторов, в таком масштабе, чтобы интервал варьирования для каждого фактора равнялся единице. В данных случаях условия проведения отдельных опытов будут соответствовать вершинам квадрата (при $k = 2$) или куба (при $k = 3$). Основной уровень варьирования фактора будет являться центром фигуры, а каждая сторона равна двум интервалам (см. рис. 3.1). Обозначения вершин квадрата и куба соответствуют номерам опытов в матрице планирования. Областью планирования экспериментов является площади, ограниченные квадратом или кубом. Аналогично располагаются экспериментальные точки при $k > 3$.

Для двух или трех факторов все возможные сочетания уровней можно найти перебором. С ростом числа факторов возникает необходимость в разработке методики построения матриц.

Традиционно используют три приема, основанных на переходе от матриц с меньшей размерностью к матрицам с большей.

Первый прием основан на следующем условии: при введении нового фактора каждая комбинация уровней исходного фактора встречается дважды — в сочетании с верхним и нижним уровнями нового фактора. Практический переход к новой матрице заключается в записи исходного плана для одного уровня нового фактора, а затем в повторении его для другого уровня. Данный прием не ограничивается размерностью матрицы.

Таблица 3.3

Матрица планирования эксперимента 2^3

№ опыта	x_1	x_2	x_3	y
1	—	+	—	y_1
2	+	—	—	y_1
3	—	—	+	y_3
4	+	+	+	y_4
5	—	+	—	y_5
6	+	—	—	y_6
7	—	—	+	y_7
8	+	+	+	y_8

Второй прием основан на правиле перемножения столбцов матрицы. Построчное перемножение уровней исходной матрицы приводит к получению дополнительного столбца произведения $x_1 \cdot x_2$. После этого повторяется исходный план, а у столбца произведений меняются знаки. Данный прием также не ограничивается размером матрицы, но в исполнении более сложный, чем первый.

Третий прием реализуется путем чередования знаков. В первом столбце знаки меняются через строку, во втором столбце — через две, в третьем — через четыре, в четвертом — через восемь и т.д. Матрица планирования 2^3 приведена в табл. 3.3.

3.3. Свойства полного факторного плана типа 2^k

Применение полного факторного эксперимента наиболее эффективно при построении линейных моделей за счет следующих свойств:

Первое свойство — алгебраическая сумма элементов вектора-столбца каждого фактора равна нулю, или

$$\sum_{j=1}^N x_{ij} = 0, \quad (3.1)$$

где $i = 1, 2, \dots, k$ — номер фактора; N — число опытов.

Именно первое свойство обеспечивает симметричность полученных моделей относительно центра эксперимента.

Второе свойство — сумма квадратов элементов каждого столбца равна числу опытов, или

$$\sum_{j=1}^N x_{ij}^2 = N. \quad (3.2)$$

Второе свойство обусловлено тем, что значения факторов в матрице задаются в виде $+1$ и -1 .

Третье свойство — сумма почленных произведений любых двух векторов-столбцов матрицы равна нулю, или

$$\sum_{j=1}^N x_{ij} \cdot x_{ju} = 0$$

при $i \neq u$, а также $i, u = 0, 1, \dots, k$. Это важнейшее свойство называется *ортogonalностью* матрицы планирования.

Четвертое свойство называется ротатабельностью, оно заключается в том, что за счет подбора точек в матрице планирования точность определения значения параметра оптимизации по построенной модели одинакова на одинаковых расстояниях от центра эксперимента и не зависит от направления.

Обеспечение вышеперечисленных условий обеспечивает минимальную дисперсию коэффициентов регрессии, а также равенство дисперсии, что облегчает статистический анализ результатов эксперимента.

3.4. Расчет коэффициентов регрессии

После построения матрицы планирования проводится эксперимент. На основании полученных экспериментальных данных рассчитываются значения коэффициентов регрессии.

Значение свободного члена b_0 определяется по следующей формуле:

$$b_0 = \frac{\sum_{j=1}^N y_j}{N},$$

где y_j — значения параметра оптимизации в j -м опыте;

N — число опытов в матрице.

Линейные коэффициенты регрессии рассчитывают по формуле

$$b_i = \frac{\sum_{j=1}^N x_{ij} y_j}{\sum_{j=1}^N x_{ij}^2} = \frac{\sum_{j=1}^N x_{ij} y_j}{N},$$

где x_{ij} — кодированное значение фактора x_i в j -м опыте.

Коэффициенты регрессии, характеризующие совместное влияние факторов на параметры оптимизации, рассчитывают по следующей формуле:

$$b_{ij} = \frac{\sum_{j=1}^N x_{ij} x_{ij} y_j}{\sum_{j=1}^N x_{ij}^2}.$$

Рассмотрим пример расчета коэффициентов регрессии для планирования 2^2 , матрица планирования которой приведена в табл. 3.2:

$$b_0 = \frac{y_1 + y_2 + y_3 + y_4}{4};$$

$$b_1 = \frac{+y_1 - y_2 + y_3 - y_4}{4};$$

$$b_2 = \frac{+y_1 + y_2 - y_3 - y_4}{4};$$

$$b_3 = \frac{+y_1 - y_2 - y_3 + y_4}{4}.$$

Рассмотрим уравнение регрессии для $k = 3$:

$$y = b_0 + b_1x + b_2x^2 + b_3x^3 + b_{12}x_1x_2 + b_{13}x_1x_3 + b_{23}x_2x_3 + b_{123}x_{123},$$

где b_0 — свободный член;

b_1, b_2, b_3 — линейные коэффициенты;

b_{12}, b_{13}, b_{23} — коэффициенты двойного взаимодействия;

b_{123} — коэффициент тройного взаимодействия.

Общее число всех коэффициентов регрессии равно числу опытов ПФП. Для того чтобы найти число взаимодействий любого порядка, можно использовать формулу числа сочетаний:

$$C_k^m = \frac{k!}{m!(k-m)!},$$

где k — число факторов;

m — число элементов во взаимодействии.

В общем случае при реализации ПФП 2^k , содержащего $N = 2^k$ опытов, число парных взаимодействий, которые могут иметь место, равно числу сочетаний из k по 2:

$$C_k^2 = \frac{k(k-1)}{2}. \quad (3.3)$$

Число взаимодействий второго порядка равно числу сочетаний из k по 2:

$$C_k^3 = \frac{k(k-1)(k-2)(k-3)}{2 \cdot 3}.$$

Так, по результатам ПФП 2^4 можно оценить:

- свободный член;
- четыре линейных члена;
- шесть парных взаимодействий ($C_4^2 = 4(4-1)/2 = 6$);
- четыре взаимодействия второго порядка ($C_4^2 = 4(4-1)(4-2) \times (4-3)/2 \cdot 3 = 4$);
- одно взаимодействие третьего порядка.

Всего $p = 1 + 4 + 6 + 4 + 1 = 16$ коэффициентов регрессии, т.е. столько, сколько поставлено опытов (насыщенное планирование).

3.5. Статистический анализ регрессионной модели, полученной по результатам ПФП 2^k

Свойства матрицы ПФП — формулы (3.1), (3.2) — упрощают не только расчет коэффициентов регрессии, но и статистический анализ регрессионной модели, полученной по результатам его реализации:

- Коэффициенты регрессии определяются независимо друг от друга и в результате отпадает необходимость в пересчете их величин после отбрасывания членов в уравнении регрессии.

- Дисперсии всех коэффициентов регрессии равны между собой и определяются по формулам:

— при равномерном дублировании опытов

$$s^2\{b_i\} = \frac{s^2\{y\}}{nN},$$

где $s^2\{y\}$ — оценка дисперсии воспроизводимости;

N — число основных опытов, запланированных матрицей, т.е. число серий опытов, каждая из которых содержит n дублированных опытов;

— при отсутствии дублированных опытов

$$s^2\{b_i\} = \frac{s^2\{y\}}{N}.$$

• Для оценки значимости коэффициентов регрессии, найденных по результатам ПФП, используется, как и в общем случае, t -критерий Стьюдента.

3.6. Применение ПФП 2³

Прочность бетона зависит от различных факторов: плотности, активности цемента, прочности заполнителя, соотношения крупного и мелкого заполнителя, технологии уплотнения бетонной смеси и т.д.

Как известно, на прочность бетона при сжатии влияют следующие основные факторы:

- водоцементное отношение — В/Ц;
- количество песка на 1 м³ бетонной смеси — П, кг;
- активность цемента — $R_{ц}$, МПа.

Пример 3.1. В эксперименте по определению прочности бетона от В/Ц, количества песка и активности цемента использовались образцы стандартного размера 100×100×100 мм. Определение прочности образцов выполнялось на гидравлическом прессе.

Требуется получить зависимость прочности бетона от следующих факторов: В/Ц (X_1), количества песка П (X_2) и активности цемента $R_{ц}$ (X_3).

Решение

Для установления этой зависимости запланирована реализация ПФП 2³.

Уровни и интервалы варьирования факторов определялись на основе предварительных экспериментов.

Значения уровней и интервалов варьирования факторов приведены в табл. 3.4. Принято число факторов $k = 3$, число уровней $p = 2$, число опытов $N = 8$, число повторных опытов $n = 3$. Матрица планирования эксперимента в нормализованных и натуральных обозначениях приведена в табл. 3.5. Необходимое число параллельных опытов n было равно трем. В табл. 3.5 также приведено среднее значение параметра оптимизации по результатам всех параллельных испытаний.

Уровни и интервалы варьирования факторов

Фактор			Уровни варьирования фактора			Интервал варьирования	Расчет фактора
Наименование	Обозначение		+1	0	-1		
	натуральное	кодированное					
Водоцементное отношение В/Ц	X_1	x_1	0,5	0,45	0,4	0,05	$x_1 = (В/Ц - 0,45)/0,05$
Количество песка П, кг	X_2	x_2	500	450	400	50	$x_2 = (П - 450)/50$
Активность цемента $R_{ц}$, МПа	X_3	x_3	60,0	50,0	40,0	10,0	$x_3 = (R_{ц} - 500)/100$

Таблица 3.5

Матрица планирования эксперимента

№ опыта	Матрица			Натуральные значения переменных			Выходная величина
	x_1	x_2	x_3	В/Ц	П	$R_{ц}$	R_6 , МПа
1	-1	-1	-1	0,4	400	40,0	31,3
2	-1	1	-1	0,4	500	40,0	40,6
3	1	-1	-1	0,5	400	40,0	22,9
4	1	1	-1	0,5	500	40,0	32,4
5	-1	-1	1	0,4	400	60,0	47,4
6	-1	1	1	0,4	500	60,0	53,9
7	1	-1	1	0,5	400	60,0	33,9
8	1	1	1	0,5	500	60,0	41,6
Σ	0	0	0	3,6	3600	400,0	304

Результаты восьми серий параллельных опытов приведены в столбцах 5—7 табл. 3.6.

В столбце 8 приведены значения отклика, усредненные по каждой из этих серий:

$$\bar{y}_j = \frac{\left(\sum_{u=1}^5 y_{ju} \right)}{3}, \quad j = 1, 2, 3, \dots, 8.$$

Результаты эксперимента

№ опыта	Значения факторов			Результаты эксперимента R_6 , МПа			Результаты расчетов		
	В/Ц	П	$R_{ц}$	y_{j1}	y_{j2}	y_{j3}	\bar{y}	s_j^2	\hat{y}
1	2	3	4	5	6	7	8	9	10
1	0,4	400	40,0	32,3	30,2	31,4	31,3	1,11	31,25
2	0,4	500	40,0	39,7	41,3	40,7	40,6	0,66	40,65
3	0,5	400	40,0	23,1	22,8	22,9	22,9	0,025	22,95
4	0,5	500	40,0	31,2	33,4	32,7	32,4	1,25	32,35
5	0,4	400	60,0	46,4	47,7	48,0	47,4	0,725	47,1
6	0,4	500	60,0	54,7	53,3	53,6	53,9	0,545	54,2
7	0,5	400	60,0	34,0	33,6	34,2	33,9	0,095	34,2
8	0,5	500	60,0	41,0	42,6	41,3	41,6	0,725	41,3
Σ								5,135	

Столбец 9 содержит результаты расчета дисперсий по каждой серии дублированных опытов:

$$s_j^2 = \frac{\left[\sum_{u=1}^5 (y_{ju} - \bar{y}_j)^2 \right]}{2}, \quad j=1, 2, 3, \dots, 8.$$

Была проведена проверка однородности дисперсий опытов. В данном эксперименте имеется равномерное дублирование, в связи с чем следует использовать критерий Кохрена. Максимальной из дисперсий оказалась дисперсия четвертого опыта $s_4^2 = 1,25$:

$$G_{\text{расч}} = \frac{s_4^2}{s_1^2 + s_2^2 + \dots + s_8^2} = \frac{1,25}{5,135} = 0,37.$$

Из таблиц распределения Кохрена для $q = 0,01, f = n - 1 = 4$ (число степеней свободы для каждой выборки), $m = 8$ (количество выборок) находим: $G_{\text{табл}} = 0,6152$.

Полученное соотношение $G_{\text{расч}} < G_{\text{табл}}$ позволяет принять гипотезу об однородности дисперсий опытов.

Находим оценку дисперсии воспроизводимости эксперимента как среднее арифметическое дисперсий опытов:

$$s^2\{y\} = \frac{\left(\sum_{j=1}^8 s_j^2\right)}{8} = \frac{5,135}{8} = 0,64.$$

Регрессионную модель объекта находим в виде неполного многочлена третьего порядка, т.е. в виде следующей формулы:

$$y = b_0 + b_1x_1 + b_2x_2 + b_3x_3 + b_{12}x_1x_2 + b_{13}x_1x_3 + b_{23}x_2x_3 + b_{123}x_1x_2x_3, \quad (3.12)$$

позволяющей учесть наряду с линейными членами все взаимодействия факторов.

Вычисление коэффициентов регрессии приведено в табл. 3.7.

В результате проведенных расчетов получены следующие коэффициенты управления регрессии: $b_0 = 38$; $b_1 = 4,125$; $b_2 = -5,3$; $b_3 = 6,2$; $b_{12} = 0,175$; $b_{13} = -0,575$; $b_{23} = -1,15$; $b_{123} = 0,125$.

Уравнение регрессии в нормализованных обозначениях факторов имеет следующий вид:

$$y = 38 + 4,125x_1 - 5,3x_2 + 6,2x_3 + 0,175x_1x_2 - 0,575x_1x_3 - 1,15x_2x_3 + 0,125x_1x_2x_3.$$

Проведем статистический анализ полученного уравнения регрессии:

- *Нахождение дисперсий коэффициентов регрессии.* Все они одинаковы и равны

$$s^2\{b_i\} = s^2\{y\}/nN;$$

$$s^2\{b_i\} = 10,64/3 \cdot 8 = 0,03.$$

- *Определение среднего квадратичного отклонения для каждого коэффициента регрессии:*

$$s\{b_i\} = \sqrt{s^2\{b_i\}} = 0,16.$$

Таблица 3.7

Вычисление коэффициентов регрессии

№	x_1	x_2	x_3	x_1x_2	x_1x_3	x_2x_3	$x_1x_2x_3$	X_1	X_2	X_3	X_1X_2	X_1X_3	X_2X_3	$X_1X_2X_3$	R_6 , МПа
1	-1	-1	-1	1	1	1	-1	-31,3	-31,3	-31,3	31,3	31,3	31,3	-31,3	31,3
2	1	-1	-1	-1	-1	1	1	40,6	-40,6	-40,6	-40,6	-40,6	40,6	40,6	40,6
3	-1	1	-1	-1	1	-1	1	-22,9	22,9	-22,9	-22,9	22,9	-22,9	22,9	22,9
4	1	1	-1	1	-1	-1	-1	32,4	32,4	-32,4	32,4	-32,4	-32,4	-32,4	32,4
5	-1	-1	1	1	-1	-1	1	-47,4	-47,4	47,4	47,4	-47,4	-47,4	47,4	47,4
6	1	-1	1	-1	1	-1	-1	53,9	-53,9	53,9	-53,9	53,9	-53,9	-53,9	53,9
7	-1	1	1	-1	-1	1	-1	-33,9	33,9	33,9	-33,9	-33,9	33,9	-33,9	33,9
8	1	1	1	1	1	1	1	41,6	41,6	41,6	41,6	41,6	41,6	41,6	41,6
Σ	0	0	0	0	0	0	0	33	-42,4	49,6	1,4	-4,6	-9,2	1,0	304

• *Оценка значимости коэффициентов регрессии.* Для каждого из коэффициентов регрессии проверим соотношение

$$|b_i| \leq t_{\text{табл}} s\{b_i\}.$$

Величина $t_{\text{табл}}$ находится из таблиц t -распределения Стьюдента при уровне значимости $q = 0,01$ и числе степеней свободы f_y . Так как имеет место равномерное дублирование, величина f_y определяется по формуле

$$f_y = N(n - 1) = 8(3 - 1) = 16;$$

$$t_{\text{табл}} = 2,92;$$

$$t_{\text{табл}} s\{b_i\} = 2,92 \cdot 0,16 = 0,47.$$

Условие незначимости коэффициентов регрессии следующее:

$$|b_i| \leq t_{\text{табл}} s\{b_i\}.$$

Следовательно,

$$b_1 = |4,125| > 0,47; \quad b_{13} = |-0,575| > 0,47;$$

$$b_2 = |-5,3| > 0,47; \quad b_{23} = |-1,15| > 0,47;$$

$$b_3 = |6,2| > 0,47; \quad b_{123} = |0,125| < 0,47.$$

$$b_{12} = |0,175| < 0,47;$$

Коэффициенты b_{12} и b_{123} признаем незначимыми, а соответствующие члены $b_{12}x_1x_2$ и $b_{123}x_1x_2x_3$ исключены из уравнения регрессии. Все остальные коэффициенты регрессии оказались значимыми.

Регрессионная модель после отбрасывания незначимых членов получена в виде

$$y = 38 + 4,125x_1 - 5,3x_2 + 6,2x_3 - 0,575x_1x_3 - 1,15x_2x_3.$$

• *Определение доверительного интервала для каждого из значимых коэффициентов.* Определение доверительного интервала производится по формуле

$$b_i - t_{\text{табл}} s\{b_i\} \leq \beta_i \leq b_i + t_{\text{табл}} s\{b_i\}.$$

В результате получены доверительные интервалы коэффициентов:

$$3,655 \leq \beta_1 \leq 4,595; \quad -1,045 \leq \beta_{13} \leq 0,105;$$

$$-5,77 \leq \beta_2 \leq -4,83; \quad -1,62 \leq \beta_{23} \leq -0,68.$$

$$5,73 \leq \beta_3 \leq 6,67;$$

• Проверка адекватности математической модели производится по следующей формуле:

$$s_{\text{ад}}^2 = S_{\text{ад}} / f_{\text{ад}} = n \sum \frac{(\bar{y}_j - \hat{y}_j)^2}{(N - p)},$$

где p — число коэффициентов регрессии оцениваемой модели;
 $s_{\text{ад}}^2 = 2,22$.

Находим значение $F_{\text{расч}}$:

$$F_{\text{расч}} = s_{\text{ад}}^2 / s^2\{y\} = 2,2/0,64 = 3,44.$$

Из таблиц F -распределения для $q = 0,01$ и чисел степеней свободы $f_{\text{ад}} = N - p = 8 - 7 = 1$ (числитель) и $f_{\text{табл}} = N(n - 1) = 32$ (знаменатель) было найдено значение $F_{\text{табл}}$, равное 3,63.

Так как $F_{\text{расч}} < F_{\text{табл}}$, была принята гипотеза об адекватности модели.

Регрессионная модель с натуральными обозначениями факторов имеет вид

$$R_6 = 560,95 + 25B/\text{Ц} - 1,26\text{Ц} - 1,023R_{\text{ц}} + 0,115B/\text{Ц} \cdot R_{\text{ц}} + 0,0023\text{Ц} \cdot R_{\text{ц}}.$$

Физический смысл полученной математической модели состоит в том, что полученное соотношение показывает взаимосвязь прочности бетона на сжатие с такими факторами, как В/Ц, количество цемента и активность цемента.

На параметр оптимизации перечисленные факторы влияют пропорционально, на что указывают линейные эффекты. Наибольшее влияние оказывает В/Ц отношение и количество цемента. Наименьшее влияние оказывают парные взаимодействия.

Для представления о влиянии факторов и взаимодействиях на отклик построим графики зависимостей. Определим влияние на отклик фактора $R_{\text{ц}}$ и взаимодействия В/Ц $R_{\text{ц}}$. Рассмотрим две прямые $R_6 = f(R_{\text{ц}})$:

• $x_1 = x_2 = 0$. Факторы В/Ц и Ц фиксируются на основном уровне.

Получим $R_6 = 38 + 6,23$ (прямая 1 на рис. 3.2);

• $x_1 = 0, x_2 = +1$. Фактор Ц фиксируется на верхнем уровне.

Будем иметь $R_6 = 43,3 + 5,05$ (прямая 2 на рис. 3.2).

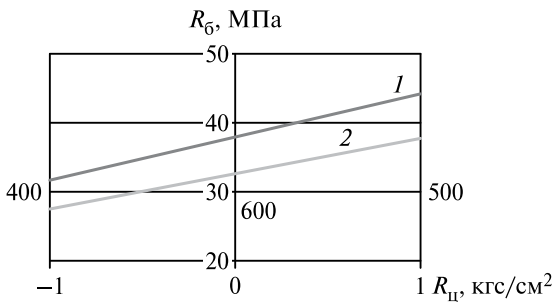


Рис. 3.2. Графики зависимостей $R_b = f(R_c)$

Поскольку с ростом фактора Ц отклик падает ($b_3 < 0$), то прямая 2 лежит ниже прямой 1. Эти прямые не параллельны. Так как $b_{23} > 0$, прямая 2 наклонена под большим углом к оси абсцисс, чем прямая 1. Из этого следует, что при большем изменении В/Ц отношения изменение активности цемента R_c оказывает более значимое влияние на прочность бетона.

3.7. Дробные факторные планы

Эксперименты в промышленности строительных материалов достаточно трудоемкие и дорогие, в связи с чем сокращение затрат на проведение эксперимента и, в частности, уменьшение числа опытов является актуальным. Решение этой проблемы возможно при применении дробных факторных планов (ДФП). При заданном числе факторов ДФП содержат меньшее число опытов по сравнению с ПФП 2^k . Это достигается упрощением математической модели. По результатам ПФП 2^k , как известно, можно оценить свободный член в регрессионной модели, линейные коэффициенты регрессии и взаимодействие факторов. Надо отметить, что учет всех взаимодействий факторов не является необходимостью.

При первоначальном изучении объектов эксперименты, как правило, ставятся с целью получения линейной модели. Для k варьируемых факторов такая модель имеет вид $y = b_0 + b_1x_1 + b_2x_2 + \dots + b_kx_k$ и содержит $k + 1$ коэффициент регрессии. Эксперимент, позволяющий найти эти коэффициенты, должен содержать не менее $k + 1$ опыт. С целью экономии средств желательно, чтобы число опытов N не слишком превышало эту величину. ПФП при отыскании линейной модели неудовлетворительны, число опытов

ПФП, равное 2^k , превосходит величину $k + 1$, начиная уже с трех факторов. ПФП неэкономичны, даже если экспериментатора интересуют, кроме линейных коэффициентов регрессии, взаимодействия факторов. Так, по результатам эксперимента с шестью факторами ($k = 6$) необходимо оценить свободный член, линейные коэффициенты регрессии и парные взаимодействия. Минимальное число опытов, необходимое для этого, равно числу коэффициентов регрессии: $N_{\min} = p = 1 + k + C_k^2 = 1 + 6 + (6 \cdot 5) / 2 = 22$, тогда как ПФП для шести факторов содержит $N = 2^6 = 64$ опыта.

ДФП позволяют сократить число опытов по сравнению с ПФП в случае, если в уравнении регрессии можно заранее пренебречь некоторыми или всеми взаимодействиями факторов. Рассмотрим построение ДФП. Обратимся сначала к ПФП 2^6 . В табл. 3.8 и 3.9 воспроизведена матрица базисных функций этого плана, соответствующая модели $y = b_0 + b_1x_1 + b_2x_2 + b_{12}x_1x_2$.

Таблица 3.8

Матрица двухфакторного эксперимента

№ опыта	x_0	x_1	x_2	$x_3(x_1x_2)$
1	+1	-1	-1	+1
2	+1	+1	-1	+1
3	+1	+1	+1	+1
4	+1	+1	+1	+1

Таблица 3.9

Матрица трехфакторного эксперимента

№ опыта	x_1	x_2	x_3
1	+1	-1	-1
2	+1	+1	-1
3	+1	+1	+1
4	+1	+1	+1

Как показывает практика, многие взаимодействия факторов незначительно влияют на выходную величину, а иногда и вовсе не влияют на нее. Таким образом, некоторое взаимодействие факторов можно заменить новым фактором, например $x_3 = x_1x_2$. Тогда для эксперимента с тремя факторами будет получен план, содержащий четыре опыта, тогда как ПФП 2^3 с тремя факторами будет содержать $N = 8$ опытов. Планы, полученные из ПФП заменой некоторых взаимодействий новыми факторами, называются дробными факторными планами. Обозначают их ДФП 2^{3-1} . План 2^{k-1} называется полурепликой от полного факторного плана.

В ДФП 2^{3-1} фактор x_3 варьируется одинаково с парным взаимодействием x_1x_2 . Обозначим через β истинные величины соответствующих коэффициентов регрессии, тогда коэффициент b_3 дает совместную или смешенную оценку двух истинных коэффициентов β_3 и β_{12} ; записывается это так: $b_3 \rightarrow \beta_3 + \beta_{12}$.

Столбцы x_1x_3 и x_2x_3 будут совпадать со столбцами x_2 и x_3 , соответственно, следовательно, будут смешанные оценки

$$b_2 \rightarrow \beta_2 + \beta_{13};$$

$$b_1 \rightarrow \beta_1 + \beta_{23}.$$

При построении ДФП 2^{3-1} использовалось соотношение $x_3 = x_1x_2$, называемое генератором плана или генерирующим соотношением.

Генерирующим соотношением или *генератором плана* называется соотношение, в котором в левой части стоит новый фактор, а в правой — произведение некоторых взаимодействий факторов.

Соотношение, в левой части которого стоит 1, а в правой — некоторое произведение факторов, называется *определяющим контрастом* реплики. При умножении обеих частей генератора плана $x_3 = x_1x_2$ на x_3 получим $(x_3)^2 = x_1x_2x_3$. С помощью *определяющего контраста* можно получить систему смешивания факторов.

Рассмотрим построение дробных реплик на основе ПФП 2^3 .

Для ПФП 2^3 модель имеет вид

$$y = b_0 + b_1x_1 + b_2x_2 + b_3x_3 + b_{12}x_1x_2 + b_{13}x_1x_3 + b_{23}x_2x_3 + b_{123}x_1x_2x_3.$$

Матрица базисных функций ПФП 2^3 для этой модели имеет вид, приведенный в табл. 3.10.

Таблица 3.10

Матрица базисных функций ПФП 2^3

№ опыта	x_0	x_1	x_2	x_3	x_1x_2	x_1x_3	x_2x_3	$x_1x_2x_3$
1	+1	-1	-1	-1	+1	+1	+1	-1
2	+1	+1	-1	-1	-1	-1	+1	+1
3	+1	-1	+1	-1	-1	+1	-1	+1
4	+1	+1	+1	-1	+1	-1	-1	-1
5	+1	-1	-1	+1	-1	-1	-1	+1
6	+1	+1	-1	+1	-1	+1	-1	-1
7	+1	-1	+1	+1	+1	-1	+1	-1
8	+1	+1	+1	+1	+1	+1	+1	+1

Пренебрегая тройным взаимодействием $x_1x_2x_3$, заменим соответствующий столбец фактором x_4 , получим план для четырех факторов (табл. 3.11).

Преобразованная матрица базисных функций ПФП 2³

№ опыта	x_1	x_2	x_3	x_4
1	-1	-1	-1	-1
2	+1	-1	-1	+1
3	-1	+1	-1	+1
4	+1	+1	-1	-1
5	-1	-1	+1	+1
6	+1	-1	+1	-1
7	-1	+1	+1	-1
8	+1	+1	+1	+1

Генератор этого плана $x_4 = x_1x_2x_3$, определяющий контраст $I = x_1x_2x_3x_4$. Умножением генератора на $x_1, x_2, x_3, x_1x_2, x_2x_3, x_1x_4$ получим новые генераторы:

$$x_1 = x_2x_3x_4; x_2 = x_1x_3x_4; x_3 = x_1x_2x_4;$$

$$x_{12} = x_3x_4; x_1x_3 = x_2x_4; x_1x_4 = x_2x_3.$$

Получаем систему смешивания оценок:

$$b_1 = \beta_1 + \beta_{234}; b_2 = \beta_2 + \beta_{134}; b_3 = \beta_3 + \beta_{124}; b_4 = \beta_4 + \beta_{123};$$

$$b_{12} = \beta_{12} + \beta_{34}; b_{13} = \beta_{13} + \beta_{24}; b_{14} = \beta_{14} + \beta_{23}.$$

Уравнение регрессии, построенное по результатам этого плана, содержит свободный член, линейные члены и парные взаимодействия, т.е.

$$y = b_0 + b_1x_1 + b_2x_2 + b_3x_3 + b_4x_4 + b_{12}x_1x_2 + b_{13}x_1x_3 + b_{23}x_2x_3. \quad (3.13)$$

По системе смешивания коэффициент b_{12} оценивает как β_{12} , так и β_{34} и т.д. план, используемый для построения модели (3.3), является насыщенным, число опытов равно числу оцениваемых коэффициентов регрессии. Проверить адекватность моде не представляется возможным.

Существуют другие способы построения дробных реплик на основе ПФП² [7].

Графическое изображение ПФП 2^3 и ДФП 2^{3-1} в факторном пространстве (для трех факторов — трехмерное пространство) представлено на рис. 3.3. ПФП 2^3 представлен кубом с восемью узлами (точками плана), а возможные ДФП 2^{3-1} — проекциями этого куба на три плоскости. То есть из восьми узлов выбираются четыре (рис. 3.3, а). Из куба можно также выбрать четыре точки из восьми, не лежащие в одной плоскости, и сформировать ДФП 2^{3-1} (рис. 3.3, б).

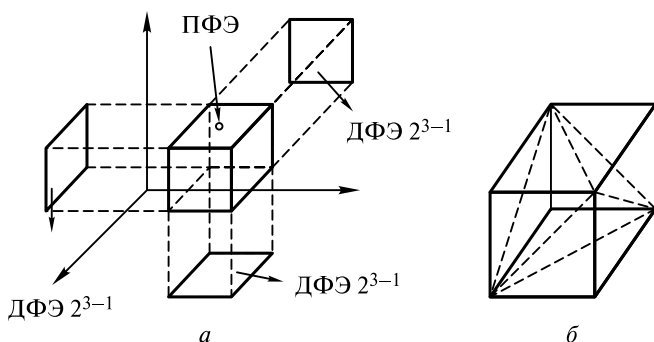


Рис. 3.3. Графическое изображение планов ПФП 2^3 и ДФП 2^{3-1} в факторном пространстве

ДФП, как и ПФП, являются ротатабельными. ДФП могут быть как насыщенными, так и ненасыщенными.

Достоинство ДФП заключается и в том, что если построенный на его основе неполный полином не удовлетворяет требованиям по точности, то план ДФП легко достраивается до плана ПФП без потери информации о прежних опытах, с формированием более точного полинома.

Глава 4. ЭКСПЕРИМЕНТАЛЬНЫЕ ПЛАНЫ ВТОРОГО ПОРЯДКА

4.1. В-планы второго порядка

Полные и дробные факторные планы позволяют получить линейное описание зависимости отклика от каждого из варьируемых факторов. При детальном изучении большинства свойств строи-

тельных материалов такое представление оказывается слишком грубым. В этой ситуации необходимо обратиться к экспериментальным планам второго порядка.

Планами второго порядка называют такие планы многофакторного эксперимента, с помощью которых можно получить математическое описание объектов в виде полинома второго порядка.

Для трех факторов соответствующее уравнение регрессии записывается в виде

$$y = B_0 + B_1X_1 + B_2X_2 + B_3X_3 + B_{11}X_1^2 + B_{22}X_2^2 + B_{33}X_3^2 + B_{12}X_1X_2 + B_{13}X_1X_3 + B_{23}X_2X_3.$$

В общем случае, когда число варьируемых факторов равно k , модель имеет вид

$$y = b_0 + \sum_{i=1}^k b_i x_i + \sum_{i=1}^k b_{ii} x_i^2 + \sum_{i<j}^k b_{ij} x_i x_j.$$

Планы, позволяющие получить линейную модель объекта, например ПФП, называют планами первого порядка.

Регрессионные модели приведенного ниже вида называют моделями второго порядка, или квадратичными моделями. Число коэффициентов регрессии такой модели составляет

$$p = \frac{1 + 2k + k(k-1)}{2} = \frac{(k+1)(k+2)}{2}. \quad (4.1)$$

Из выражения (4.1) видно, что реализация плана второго порядка позволяет описать зависимость выходной величины от каждого фактора в виде уравнения параболы.

В B -планах второго порядка каждый фактор X_i варьируется на трех уровнях, т.е. принимает в каждом опыте одно из трех значений: наименьшее $X_{i\min}$, наибольшее $X_{i\max}$ и среднее $X_i^0 = (X_{i\min} + X_{i\max})/2$. В нормализованных обозначениях эти уровни обозначаются соответственно -1 , $+1$, 0

Назовем *звездной точкой* B -плана условия опыта, в котором один из факторов принимает нормализованное обозначение $+1$ или -1 , а остальные фиксируются на основном уровне (ноль в кодовых обозначениях).

Например, некоторые звездные точки для плана с тремя факторами (в кодовых обозначениях) имеют следующий вид:

$$x_1 = +1, x_2 = 0, x_3 = 0;$$

$$x_1 = -1, x_2 = 0, x_3 = 0;$$

$$x_1 = 0, x_2 = +1, x_3 = -1;$$

$$x_1 = 0, x_2 = -1, x_3 = +1.$$

При числе факторов k имеется $2k$ различных звездных точек.

B -план состоит из точек ПФП, к которым добавлено $2k$ звездных точек. Общее число опытов B -плана, таким образом, равно $N = 2^k + 2k$. Для $k = 5$ B -план с ПФП содержит $2^5 + 2 \cdot 5 = 42$ опыта. В табл. 4.1 и 4.2 приведены B -планы для $k = 2$ и $k = 3$.

Таблица 4.1

B -план для $k = 2$ (план B_2)

№ опыта	x_1	x_2	№ опыта	x_1	x_2
ПФП			Звездные точки		
1	-1	-1	5	-1	0
2	+1	-1	6	+1	0
3	-1	+1	7	0	-1
4	+1	+1	8	0	+1

Таблица 4.2

B -план для $k = 3$ (план B_3)

№ опыта	x_1	x_2	x_3	№ опыта	x_1	x_2	x_3
ПФП				Звездные точки			
1	-1	-1	-1	9	-1	0	0
2	+1	-1	-1	10	+1	0	0
3	-1	+1	-1	11	0	-1	-1
4	+1	+1	-1	12	0	+1	0
5	-1	-1	+1	13	0	0	-1
6	+1	-1	+1	14	0	0	+1
7	-1	+1	+1	—	—	—	—
8	+1	+1	+1	—	—	—	—

Опыты, входящие в состав B -плана, обычно называют его ортогональной частью.

Изобразим опыты B -плана для $k = 2$ на факторной плоскости. Как видно из рис. 4.1, звездные точки 5—8, дополняющие ПФП до B -плана, расположены в серединах сторон квадрата, вершинами которого являются точки ПФП. В более общем случае при $k \geq 3$ опыты ортогональной части плана образуют вершины куба (гиперкуба при $k \geq 4$) в пространстве факторов (см. рис. 4.1). Звездные точки представляют собой центры граней этого куба.

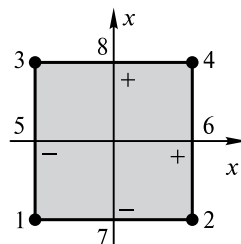


Рис. 4.1. Представление опытов плана B_2 на факторной плоскости

Из геометрического рассмотрения следует, что область варьирования факторов в B -плане (как и в ПФП) является куб (гиперкуб) в пространстве варьируемых факторов.

Все опыты B -плана расположены на границах области варьирования факторов, в области центра плана, т.е. при значениях факторов, близких к основному уровню, опытов не ставится. Из-за этого точность регрессионной модели близи центра плана оказывается недостаточно высокой. Если построенная по результатам эксперимента модель служит для оптимизации объекта, то данное обстоятельство может не удовлетворять исследователя. B -план может быть дополнен одним или несколькими опытами в центре плана, т.е. опытами, поставленными в условиях $x_1 = x_2 = \dots = x_k = 0$. Данные опыты увеличивают точность модели вблизи центра плана. Опыт, повторяющийся в центре плана, дает возможность оценить дисперсию воспроизводимости, если остальные опыты плана не дублируются. Данный план является разновидностью B -плана.

План второго порядка, содержащий в своем составе план первого порядка, называют композиционным. Свойства композиционного плана позволяют планировать и проводить эксперимент поэтапно. На первом этапе ставится план (эксперимент) первого порядка (ПФП), по результатам которого получается линейная модель или модель, содержащая, кроме линейных членов, взаимодействия факторов. Если такая модель окажется неадекватной, дополнительно ставятся опыты в звездных точках. Вся совокупность этих опытов составит B -план второго порядка. Обработка результатов такого плана позволяет получать квадратичную модель.

Расчет коэффициентов регрессии B -плана ведется по следующим формулам:

$$\left. \begin{aligned}
 b_0 &= T_1 \sum_{j=1}^N (0y) - T_2 \sum_{i=1}^k (i iy), \\
 b_i &= T_3 \sum_{j=1}^N (iy), \\
 b_{ii} &= T_4 (i iy) + T_5 \sum_{i=1}^k (i iy) - T_2 (0y), \\
 b_{iu} &= T_6 (i uy),
 \end{aligned} \right\} \quad (4.2)$$

где b_0 — свободный член; b_i — линейные коэффициенты регрессии, $i = 1, 2, \dots, k$; b_{ii} — квадратичные коэффициенты регрессии, $i = 1, 2, \dots, k$; b_{iu} — коэффициент при парных взаимодействиях, $i \neq u$; T_i — коэффициенты, значения которых указаны в табл. 4.3.

$$\left. \begin{aligned}
 (0y) &= \sum_{j=1}^N y_j, \\
 (i iy) &= \sum_{j=1}^N x_{ij}^2 y_j, \\
 (iy) &= \sum_{j=1}^N x_{ij} y_j, \\
 (i uy) &= \sum_{j=1}^N x_{ij} x_{uj} y_j, \quad i \neq u.
 \end{aligned} \right\} \quad (4.3)$$

Дисперсии коэффициентов регрессии и ковариации между ними определяются по следующим формулам:

$$\begin{aligned}
 s^2\{b_0\} &= (T_1/n) s^2\{y\}; \\
 s^2\{b_i\} &= (T_3/n) s^2\{y\}; \\
 s^2\{b_{ii}\} &= [(T_4 + T_5)/n] s^2\{y\}; \\
 s^2\{b_{ij}\} &= (T_6/n) s^2\{y\}; \\
 \text{cov}\{b_0, b_{ii}\} &= -(T_2/n) s^2\{y\}; \\
 \text{cov}\{b_0, b_{ij}\} &= (T_5/n) s^2\{y\}.
 \end{aligned} \quad (4.4)$$

Коэффициенты $T_1—T_6$ для B -планов с ПФП

T_i	Вид плана			
	B_2 ($k = 2, N = 8$)	B_3 ($k = 4, N = 14$)	B_4 ($k = 4, N = 24$)	B_5 с ПФП в ортогональной части ($k = 5, N = 42$)
T_1	1,25	0,40624	0,22917	0,15821
T_2	0,75	0,15624	0,0625	0,0332
T_3	0,16667	0,1	0,05556	0,02941
T_4	0,5	0,5	0,5	0,5
T_5	0,25	-0,09375	-0,10417	-0,0918
T_6	0,25	0,125	0,0625	0,03125

Рассмотрим два случая применимости этих формул:

- Отсутствие дублированных опытов (не считая опытов в центре плана). В этом случае N — число запланированных опытов. Для плана B_k с ПФП в ортогональной части и с n_0 опытами в центре плана $N = 2^k + 2k + n_0$; y_i — результат i -го опыта, $i = 1, 2, \dots, N$. Величина n в формулах (4.4), представляющая собой число дублированных опытов в каждой серии, принимается равной 1.

- Равномерное дублирование. Формулы (4.2), (4.3) по-прежнему справедливы, но под y_j понимается среднее арифметическое по результатам j -й серии дублированных опытов; N — число дублированных опытов.

Значения коэффициентов $T_1—T_6$ для B -планов с ПФП в ортогональной части и число факторов $k = 2—5$ при отсутствии опытов в центре плана приведены в табл. 4.3.

4.2. Применение B -плана второго порядка

4.2.1. РАСЧЕТ УРАВНЕНИЯ РЕГРЕССИИ

В качестве примера B -плана второго порядка можно рассмотреть эксперимент по исследованию зависимостей прочностных свойств сталефибробетонов, подвергаемых тепловлажностной обработке [8].

Для разработки методики проектирования составов сталефибробетонов, подвергаемых тепловлажностной обработке, было необходимо проведение исследований в два этапа. На первом этапе —

определить зависимости прочностных свойств сталефибробетонов, подвергаемых тепловлажностной обработке, от их составов, на втором — оптимизировать режимы тепловой обработки сталефибробетонов [8].

Пример 4.1. Для реализации задачи по определению зависимостей прочностных свойств сталефибробетона, подвергаемого тепловлажностной обработке, от его состава было проведено исследование на основе трехфакторного эксперимента.

При исследовании состава сталефибробетона были приняты не фактический расход какого-либо из компонентов, а их цементно-песчаное (Ц/П) и водоцементное (В/Ц) соотношения, что позволило двумя параметрами, которые возможно варьировать в пределах ограниченных правилом абсолютных объемов, характеризовать состав цементно-песчаной матрицы. Уровни варьирования Ц/П, В/Ц и коэффициента армирования сталефибробетона по объему μ_v были назначены на основе анализа результатов предыдущих исследований [8].

Требуется определить зависимость прочности сталефибробетона, подвергаемого тепловлажностной обработке, от указанных выше факторов.

Решение

Формулы, связывающие нормализованные и натуральные обозначения, будут иметь следующий вид:

$$x_1 = \frac{\text{Ц/П} - 0,67}{34}; \quad x_2 = \frac{\text{В/Ц} - 0,4}{0,1}; \quad x_3 = \frac{\mu_v - 2,5}{2,0}.$$

Факторы и уровни варьирования эксперимента приведены в табл. 4.4.

Для осуществления эксперимента был принят композиционный план на кубе типа B_3 . В процессе экспериментальных исследований первого этапа предполагалось получить ряд уравнений, отображающих характерные зависимости от принятых факторов и уровней их варьирования следующих свойств сталефибробетона: кубиковой прочности на сжатие сразу после тепловлажностной обработки в возрасте 28 суток после тепловлажностной обработки, прочности на растяжение при изгибе в возрасте 28 суток после тепловлажностной обработки:

$$y(R_{\text{сж,н.у}}, R_{\text{сж,0}}, R_{\text{сж,1}}, R_{\text{сж,28}}, R_{\text{из,н.у}}, R_{\text{из,28}}) = f(\mu_v, \text{Ц/П}, \text{В/Ц}).$$

Факторы и уровни варьирования эксперимента

Факторы	Обозначение		Интервал варьирования	Уровни варьирования		
	натуральное	нормализованное		Нижний (-1)	Основной (0)	Верхний (+1)
Цементно-песчаное отношение	Ц/П	x_1	0,34	0,33	0,67	1
Водоцементное отношение	В/Ц	x_2	0,1	0,3	0,4	0,5
Объемное содержание фибры в бетоне, %	μ_v	x_3	2,0	0,5	2,5	4,5

Матрица планирования и результаты проведенных испытаний для шести выбранных параметров оптимизации приведены в табл. 4.5. y_7 и y_8 являются расчетными параметрами оптимизации и определяются как разность между $R_{сж,н.у}$ и $R_{из,н.у}$ и $R_{сж,28}$ и $R_{из,28}$. При этом, хотя данные параметры оптимизации и определяются расчетным путем из полученных ранее значений, для них также могут быть построены уравнения регрессии и найдены свои экстремальные значения.

В табл. 4.6 и 4.7 приведены результаты расчета уравнения регрессии для параметра оптимизации y_3 (прочности при сжатии сталефибробетона при твердении в нормальных условиях в возрасте 28 суток).

В пятом столбце табл. 4.6 приведены значения прочности сталефибробетона в возрасте 28 суток, усредненные по трем дублированным опытам каждой серии:

$$\bar{y}_j = \frac{\sum_{u=1}^3 y_{ju}}{3}.$$

По критерию Кохрена проверяется однородность дисперсий опытов:

$$G_{расч} = 0,1327 \quad G_{табл} = 0,35 \quad \text{для } q = 0,05.$$

Матрица планирования эксперимента

№ опыта	Ц/П	В/Ц	η, %	Кодированные значения факторов			Параметры оптимизации							
				x ₁	x ₂	x ₃	y ₁ R _{СЖ,0,9} , МПа	y ₂ R _{СЖ,1,9} , МПа	y ₃ R _{СЖ,28,9} , МПа	y ₄ R _{СЖ,1,9} , МПа	y ₅ R _{ИЗ,28,9} , МПа	y ₆ R _{ИЗ,1,9} , МПа	y ₇ ΔR _{СЖ} , МПа	y ₈ ΔR _{ИЗ} , МПа
				1	2	3	4	5	6	7	8	9	10	11
1	1:3	0,3	0,5	-1	-1	-1	16,9	22,6	32,8	27,1	4,0	3,8	5,7	0,2
2	1:1	0,3	0,5	+1	-1	-1	42,2	49,1	68,2	69,3	8,1	8,2	-1,1	-0,1
3	1:3	0,5	0,5	-1	+1	-1	12,7	16,2	19	17,8	2,4	2,6	1,2	-0,2
4	1:1	0,5	0,5	+1	+1	-1	21,9	25,9	47,1	49	5,7	5,6	-1,9	0,1
5	1:3	0,3	4,5	-1	-1	+1	23,3	33,6	48	38	16,4	13,2	10	3,2
6	1:1	0,3	4,5	+1	-1	+1	49,8	57,6	72	79,7	24,6	26,6	-7,7	-2
7	1:3	0,5	4,5	-1	+1	+1	22,8	31,4	34,1	26,2	11,1	9,9	7,9	1,2
8	1:1	0,5	4,5	+1	+1	+1	30,1	34,4	53,7	61	18,2	19,6	-7,3	-1,4
9	1:3	0,4	2,5	-1	0	0	17,5	23,4	30	28,7	8,1	7,8	1,3	0,3
10	1:1	0,4	2,5	+1	0	0	34,5	37,8	47,3	56,4	12,3	13,2	-9,1	-0,9
11	1:1,5	0,3	2,5	0	-1	0	33	42,5	62,5	52,3	15,9	17,1	10,2	-1,2
12	1:1,5	0,5	2,5	0	+1	0	24,8	32,1	41,2	38	10,5	11,3	3,2	-0,8
13	1:1,5	0,4	0,5	0	0	-1	30,8	38,2	52,3	45,8	6,1	5,2	6,5	0,9
14	1:1,5	0,4	4,5	0	0	+1	24,8	37,7	62,8	54,9	19,9	17,9	7,9	2

Расчет коэффициентов уравнения регрессии

Точки плана	Факторы			Средне-арифметическое значение прочности в возрасте 28 суток, \bar{y}_3	Расчетные параметры для определения коэффициентов при:											
	x_1	x_2	x_3		линейных членах				квадратичных членах				взаимодействий			
					$\bar{x}_1\bar{y}_3$	$\bar{x}_2\bar{y}_3$	$\bar{x}_3\bar{y}_3$	$(x_1)^2\bar{y}_3$	$(x_2)^2\bar{y}_3$	$(x_3)^2\bar{y}_3$	$x_1\bar{x}_2\bar{y}_3$	$x_1\bar{x}_3\bar{y}_3$	$x_2\bar{x}_3\bar{y}_3$	$x_1x_2x_3\bar{y}_3$		
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15		
1	-1	-1	-1	32,8	-32,8	-32,8	-32,8	32,8	32,8	32,8	32,8	32,8	32,8	-32,8		
2	+1	-1	-1	68,2	68,2	-68,2	-68,2	68,2	68,2	68,2	-68,2	-68,2	68,2	68,2		
3	-1	+1	-1	19,0	-19,0	19,0	-19,0	19,0	19,0	19,0	-19,0	19,0	-19,0	19,0		
4	+1	+1	-1	47,1	47,1	47,1	-47,1	47,1	47,1	47,1	47,1	-47,1	-47,1	-47,1		
5	-1	-1	+1	48,0	-48,0	-48,0	48,0	48,0	48,0	48,0	48,0	-48,0	-48,0	48,0		
6	+1	-1	+1	72,0	72,0	-72,0	72,0	72,0	72,0	72,0	-72,0	72,0	-72,0	-72,0		
7	-1	+1	+1	34,1	-34,1	34,1	34,1	34,1	34,1	34,1	-34,1	-34,1	34,1	-34,1		
8	+1	+1	+1	53,7	53,7	53,7	53,7	53,7	53,7	53,7	53,7	53,7	53,7	53,7		
9	-1	0	0	30,0	-30,0	0,0	0,0	30,0	0,0	0,0	0,0	0,0	0,0	0,0		
10	+1	0	0	47,3	47,3	0,0	0,0	47,3	0,0	0,0	0,0	0,0	0,0	0,0		
11	0	-1	0	62,5	0,0	-62,5	0,0	0,0	62,5	0,0	0,0	0,0	0,0	0,0		
12	0	+1	0	41,2	0,0	41,2	0,0	0,0	41,2	0,0	0,0	0,0	0,0	0,0		
13	0	0	-1	52,3	0,0	0,0	-52,3	0,0	0,0	52,3	0,0	0,0	0,0	0,0		
14	0	0	+1	62,8	0,0	0,0	62,8	0,0	0,0	62,8	0,0	0,0	0,0	0,0		
	Σ			671,0	124,4	-88,4	51,2	452,2	478,6	490,0	-11,7	-19,9	2,7	2,9		

Оценка воспроизводимости полученных результатов

Номер опыта	Результаты дублированных опытов y_{ji}				Среднеарифметическое значение прочности в возрасте 28 суток, \bar{y}_3	$(y_{31} - \bar{y}_3)^2$	$(y_{31} - \bar{y}_3)^2$	$(y_{31} - \bar{y}_3)^2$	Дисперсия опыта s^2
	y_{31}	y_{32}	y_{33}						
1	38,6	34,4	25,4	32,8	33,64	2,56	54,76	45,48	
2	64,4	68,1	72,1	68,2	14,44	0,01	15,21	14,83	
3	15,5	23,6	17,9	19,0	12,25	21,16	1,21	17,31	
4	43,5	50,7	47,1	47,1	12,96	12,96	0,00	12,96	
5	52,9	43,7	47,4	48,0	24,01	18,49	0,36	21,43	
6	70,6	67,2	78,2	72,0	1,96	23,04	38,44	31,72	
7	38,5	33,1	30,7	34,1	19,36	1,00	11,56	15,96	
8	49,5	52,8	58,7	53,7	17,36	0,75	25,33	21,72	
9	28,7	24,9	36,4	30,0	1,69	26,01	40,96	34,33	
10	42,7	51,3	47,9	47,3	21,16	16,00	0,36	18,76	
11	67,1	57,6	62,8	62,5	21,16	24,01	0,09	22,63	
12	40,2	36,3	47,2	41,2	1,07	24,34	35,60	30,50	
13	47,2	57,4	52,3	52,3	26,01	26,01	0,00	26,01	
14	68,9	58,7	60,8	62,8	37,21	16,81	4,00	29,01	
			Σ	671,0	244,28	213,15	227,89	342,66	

Это позволяет оценить дисперсию воспроизводимости $s^2\{y\}$ как среднее арифметическое дисперсий опытов:

$$s^2\{y\} = \sum_{j=1}^{14} s_j^2 / 14; \quad s^2\{y\} = 342,66 / 14 = 24,48.$$

Связанное с ней число степеней свободы f_y равно

$$f_y = N(n - 1) = 28.$$

На основании расчетов по табл. 4.6, 4.7 даны оценки коэффициентов регрессии:

b_0	b_1	b_2	b_3	b_{11}	b_{22}	b_{33}	b_{12}	b_{13}	b_{23}	b_{123}
50,62	12,44	-8,85	5,12	-11,95	1,27	6,96	-1,47	-2,49	0,34	0,04

Оценки дисперсий коэффициентов регрессии определяем по формулам подразд. 2.3.3.

$$\begin{aligned} s^2\{b_0\} &= 3,314 \rightarrow s\{b_0\} = 1,820 & t_{\text{табл}} s^2\{b_0\} &= 3,731; \\ s^2\{b_i\} &= 1,274 \rightarrow s\{b_i\} = 1,129 & t_{\text{табл}} s^2\{b_i\} &= 2,314; \\ s^2\{b_{ii}\} &= 3,3143 \rightarrow s\{b_{ii}\} = 1,820 & t_{\text{табл}} s^2\{b_{ii}\} &= 3,732; \\ s^2\{b_{ij}\} &= 1,0198 \rightarrow s\{b_{ij}\} = 1,009 & t_{\text{табл}} s^2\{b_{ij}\} &= 2,070. \end{aligned}$$

Согласно методике оценки значимости коэффициентов регрессии определено табличное значение t -критерия Стьюдента для числа степеней свободы $f_y = 28$. Из таблиц t -критерия (табл. П2 приложения) для $q = 0,05$ найдено $t_{\text{табл}} = 2,05$. Для коэффициента b_0 имеем

$$\begin{aligned} s\{b_0\} &= \sqrt{s^2\{b_0\}} = \sqrt{3,314} = 1,820; \\ t_{\text{табл}} s\{b_0\} &= 2,02 \cdot 1,820 = 3,731. \end{aligned}$$

Для коэффициента b_0 соотношение (2.12) из подразд. 2.3.3 не выполняется, следовательно, он значим. Аналогично установлена зависимость всех коэффициентов регрессии.

Получено следующее уравнение регрессии:

$$\begin{aligned} y_3 &= 50,61 + 12,44x_1 - 8,84x_2 + 5,12x_3 - \\ &\quad - 11,95(x_1)^2 + 6,95(x_3)^2 - 2,49x_1x_3. \end{aligned}$$

Проверка адекватности уравнения регрессии проводится в следующей последовательности:

- По полученному уравнению регрессии вычислить значения отклика \hat{y} для каждого опыта.
- Определить дисперсию адекватности по формуле (2.14) из подразд. 2.3.3: $s_{ад}^2 = 39,27$.
- Вычислить расчетное F -отношение: $F_{расч} = s_{ад}^2 / s^2\{y\} = 1,61$.
- Расчетное значение F -отношения сравнить с табличным отношением $F_{табл}$ (табл. П4 приложения) для уровня значимости q и чисел степеней свободы $f_{ад}$ (числитель) и f_y (знаменатель):

$$f_{ад} = N - p = 14 - 7 = 7,$$

где N — число опытов;

p — число коэффициентов регрессии.

Для $q = 0,05$ $F_{табл} = 2,36$. Полученное соотношение $F_{расч} < F_{табл}$ позволяет принять гипотезу об адекватности регрессионной модели.

Конечным результатом обработки данных эксперимента явилось получение уравнений регрессий по количеству параметров оптимизации со значимыми коэффициентами при выбранных переменных:

$$R_{сж0} = 27,71 + 6,98(X_1) - 5,76(X_2) + 3,1(X_3) - 6,75(X_1)(X_2); \quad (4.5)$$

$$R_{сж1} = 36 + 7,76(X_1) - 6,54(X_2) + 4,27(X_3) - 4,73(X_1)(X_2) - 5,4(X_1)^2; \quad (4.6)$$

$$R_{сж28} = 50,61 + 12,44(X_1) - 8,84(X_2) + 5,12(X_3) - 11,95(X_1)^2 + 6,95(X_3)^2 - 2,49(X_1)(X_3); \quad (4.7)$$

$$R_{сж.н.у} = 46,02 + 17,76(X_1) - 7,74(X_2) + 5,08(X_3) - 3,47(X_1)^2 + 4,43(X_3)^2; \quad (4.8)$$

$$R_{изг28} = 12,54 + 2,69(X_1) - 2,11(X_2) + 6,39(X_3) - 0,99(X_1)(X_3) - 0,96(X_2)(X_3) - 2,34(X_1)^2; \quad (4.9)$$

$$R_{изг.н.у} = 12,53 + 3,59(X_1) - 1,99(X_2) + 6,18(X_3) + 1,96(X_1)(X_3) - 0,81(X_2)(X_3) - 2,03(X_1)^2 + 1,67(X_2)^2 - 0,98(X_3)^2. \quad (4.10)$$

Уравнения регрессии (4.5)—(4.10) получены аналогично уравнению (4.8), подробный расчет которого приведен ранее.

4.2.2. АНАЛИЗ И ИНТЕРПРЕТАЦИЯ УРАВНЕНИЯ РЕГРЕССИИ

Если уравнение регрессии отличается от линейного, то простое сравнение по абсолютной величине линейных коэффициентов регрессии не определяет относительную степень влияния факторов, так как присутствуют квадратичные члены и парные взаимодействия.

В общем случае, для i -го фактора критерием максимальной степени влияния в диапазоне варьирования является максимальное по модулю значение величины, которое равно

$$|d_{i \max}| = |b_i| + 2|b_{ii}| + \sum_{j \neq i} |b_{ij}|.$$

Рассматривая уравнение параболы общего вида

$$y = b_0 + b_1 x_i + b_{ii} x_i^2,$$

вспомним:

- если $b_{ii} > 0$, то уравнение параболы описывает вогнутую функцию (ветви параболы направлены вверх), а если $b_{ii} < 0$, то выпуклую;
- абсцисса вершины параболы равна $x_{iB} = -b_i/2b_{ii}$. Отсюда вытекают остальные свойства;
- если имеет место соотношение

$$|b_i| > 2|b_{ii}|,$$

то вершина параболы находится вне диапазона варьирования фактора x_i и, следовательно, уравнение описывает монотонную функцию. Если при этом $b_i > 0$, то функция монотонно возрастает, а при $b_i < 0$ — монотонно убывает;

- при наличии соотношения $|b_i| < 2|b_{ii}|$ функция имеет экстремум внутри диапазона варьирования фактора x_i — максимум при $b_i < 0$ или минимум при $b_i > 0$.

Пример 4.2. Требуется провести анализ рассчитанного на основе данных примера 4.1 уравнения регрессии для y_3 :

$$y_3 = 50,61 + 12,44 x_2 - 8,84 x_2^2 + 5,12 x_3 - 11,95(x_1)^2 + 6,95(x_3)^2 - 2,49 x_1 x_3.$$

Решение

Зависимость y от фактора x_2 является линейной, так как квадратичный член отсутствует. Можно утверждать, что с увеличением

величины В/Ц, соответствующей нормализованному значению фактора x_2 , отклик уменьшается при любых значениях остальных факторов. Это подтверждают соотношения: $b_2 < 0$ и $b_2 < \sum |b_{2j}|$.

Зависимости отклика от факторов x_1 и x_3 описываются уравнениями парабол, так как b_{11} и b_{33} отличны от нуля.

Рассмотрим графическую зависимость y от x_1 при различных значениях x_3 и фиксированном уровне фактора $x_2 = -1$. При таком значении фактора x_2 достигается максимум прочности. При этом проявится эффект парного взаимодействия b_{13} .

Подставив значения $x_3 = 1, x_2 = -1$ в уравнение регрессии, получим, что $y = 71,52 + 9,95x_1 - 11,95x_1^2$.

Рассмотрим поведение параболы в диапазоне $-1 \leq x_2 \leq +1$. Данная парабола выпуклая и имеет максимум внутри диапазона варьирования, так как выполняются условия: $b_{ii} < 0$ и $|b_{ii}| < 2|b_{ij}|$. Максимум находится в точке $x_{1B} = -9,95/2(-11,95) = 0,4$. Подставив найденное значение $x_1 = x_{1B}$ в уравнение, найдем величину максимума $y_B = 65,61$ (рис. 4.2, кривая 1).

Оставим значение $x_3 = 1$, примем значение $x_2 = 1$ и подставим в полученное уравнение регрессии. Получили зависимость

$$y = 53,85 + 9,95x_1 - 11,95x_1^2.$$

Наблюдаем изменение свободного члена; из этого следует, что значение прочности уменьшится. Вид параболы не изменится, так как коэффициент b_1 не изменился. Не изменилось и положение вершины параболы (рис. 4.2, кривая 2). Аналогично анализируется влияние на отклик других факторов.

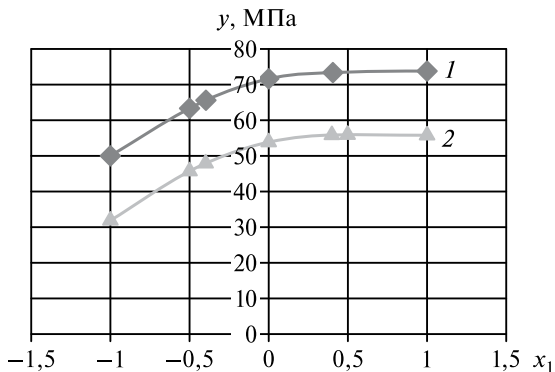


Рис. 4.2. Графики зависимостей $y = f(\text{Ц}/\text{П})$

Исследуем полученное уравнение для нахождения оптимальных условий функционирования объекта. Определим значения факторов, при которых функция отклика в модели

$$y = 50,61 + 12,44x_1 - 8,84x_2 + 5,12x_3 - 11,95(x_1)^2 + 6,95(x_3)^2 - 2,49x_1x_3,$$

т.е. прочность сталефибробетона в возрасте 28 суток будет максимальной.

Разложим уравнение на систему квазиоднофакторных моделей:

$$y_1 = 12,44x_1 - 11,95(x_1)^2 - 2,49x_1x_3; \quad (4.11)$$

$$y_2 = -8,84x_2; \quad (4.12)$$

$$y_3 = 5,12x_3^2 + 6,95(x_3)^2 - 2,49x_1x_3. \quad (4.13)$$

Для всех моделей выполняется условие $|b_i| \geq \sum_j |b_{ij}|$.

Для моделей (4.12) и (4.13) выполняется соотношение $b_{ii} \geq 0$.

Определяем оптимальные значения факторов x_2 и x_3 :

$$x_{2\text{опт}} = -1, \text{ так как } b_2 < 0, x_{3\text{опт}} = 1, \text{ так как } b_3 > 0.$$

Для модели (4.14) $b_{ii} < 0$, поэтому $x_{1\text{опт}}$ определяется как функция

$$x_{i\text{в}} = x_{i\text{опт}} = \frac{-b_i}{2b_{ii}} - \sum_{i \neq j} \frac{b_{ij}x_j}{2b_{ii}};$$

$$x_{1\text{опт}} = \frac{12,44}{2(-11,49)} - \frac{(-2,49)}{2(-11,49)} = 0,43 \approx 0,4;$$

$$y_{\text{опт}} = 50,61 + 12,44 \cdot 0,4 + 8,84 + 5,12 - 11,95(0,4)^2 + 6,95 - 2,49 \cdot 0,4 = 68,65.$$

Кроме решения задачи по определению максимального значения функции (рассчитанного уравнения регрессии) достаточно часто встречаются задачи, в которых необходимо определить усредненные значения параметра оптимизации в заданных интервалах либо определить общее направление влияния отдельного фактора. Для решения подобных задач достаточно наглядно и удобно

воспользоваться графическим отображением полученных уравнений. Для укрупненной оценки влияния различных факторов в большей степени подходит построение плоскости в координатах x, y, z , где параметр оптимизации откладывается по оси z , а факторы x_1 и x_2 — по осям x и y , соответственно. Значение фактора x_3 принимается равным одному из допустимых значений (в большинстве случаев соответствующих предельным и среднему значению). В результате в системе координат x, y, z воспроизводятся три плоскости, каждая из которых соответствует определенному постоянному значению x_3 . Характер положения данных плоскостей, наличие пиков и перепадов по высоте (оси z) и позволяет дать общую оценку влияния факторов на параметр оптимизации в рассматриваемых интервалах.

Для более точной оценки влияния факторов на параметр оптимизации используют изолинии в плоскостных координатах. В данном случае факторы x_1 и x_2 заданы по осям x и y , соответственно, фактор x_3 принимается равным любому из допустимых значений, а параметр оптимизации y изображается в данной плоскости в виде изолиний, каждой из которых соответствует определенное значение.

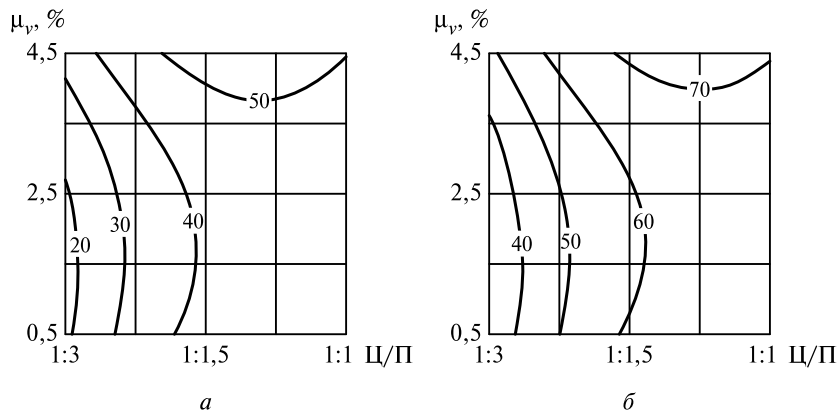


Рис. 4.3. Изолинии прочности сталефибробетона при сжатии в возрасте 28 суток после тепловлажностной обработки при $V/C = 0,5$ (а) и $V/C = 0,3$ (б)

На рис. 4.3 и 4.4 приведены примеры изолиний прочности сталефибробетона при сжатии в возрасте 28 суток после тепловлажностной обработки, полученные при анализе уравнения (4.7).

По полученным изолиниям достаточно легко, наглядно и точно можно определить как оптимальные значения факторов, так и влияние отдельных факторов на параметр оптимизации. К недостаткам данного способа анализа полученных уравнений регрессии относятся значительный объем расчетных работ, а также большое количество отдельных графиков с изолиниями, каждый из которых требует детального рассмотрения.

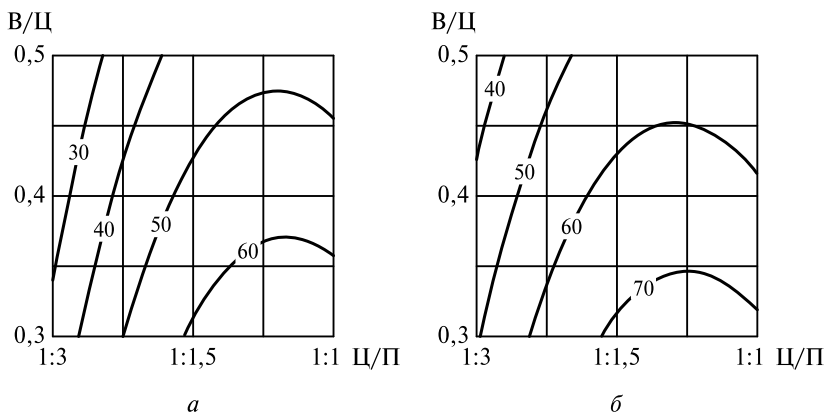


Рис. 4.4. Изолинии прочности сталефибробетона при сжатии в возрасте 28 суток после тепловлажностной обработки при $\mu_v = 0,5\%$ (а) и $\mu_v = 4,5\%$ (б)

4.3. Униформ-ротатабельный план

Униформ-ротатабельный план (УРП) относится к планам второго порядка. Этот план обладает свойствами равномерности и ротатабельности.

Свойство *ротатабельности* означает, что точность уравнения регрессии, полученного по результатам этого плана, будет одинакова во всех точках экспериментального пространства, находящихся на дном уровне от центра плана. Свойство равномерности в сочетании с ротатабельностью означает постоянство дисперсии $s^2\{\hat{y}\}$ в окрестности центра плана. Униформ-ротатабельные планы с достаточно большой точностью описывают объект вблизи центра плана и со значительно большей погрешностью — на границах области варьирования факторов.

Структура униформ-ротатабельного плана аналогична структуре В-плана. Ротатабельный план — композиционный план, он со-

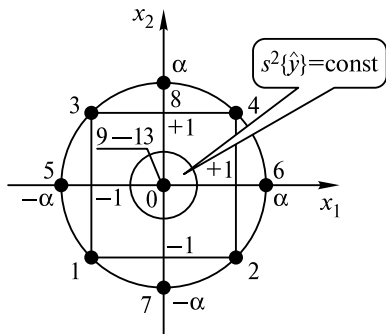


Рис. 4.5. УРП для двух факторов на факторной плоскости

держит в ортогональной части полный факторный план, к опытам которого добавлено $2k$ звездных точек и некоторое число n_0 опытов в центре плана (рис. 4.5).

Число опытов в центре УРП задается в зависимости от числа варьируемых факторов исходя из требования равномерности плана.

Опыты полного факторного плана в ортогональной части униформ-ротатабельного плана находятся внутри области варьирования факторов, а не на границах диапазона варьирования, значения

в этих опытах обозначаются -1 и $+1$, а нижний и верхний уровни каждого фактора обозначаются $-\alpha$ и $+\alpha$ (α — положительное число, больше 1, называется *звездным плечом*). Как и в B -плане, звездные точки — опыты, в которых один из варьируемых факторов фиксируется на верхнем или нижнем уровне, а остальные — на основном (нулевом) уровне:

$$x_1 = \pm\alpha, x_2 = x_3 = \dots = x_k = 0;$$

$$x_2 = \pm\alpha, x_1 = x_3 = \dots = x_k = 0, \text{ и т.п.}$$

Величина звездного плеча α определяется по формуле $\alpha = 2^{k/2}$, если в ортогональной части содержится ПФП, или по формуле $\alpha = 2^{(k-1)/4}$, если план содержит полуреплику.

В табл. 4.8 и 4.9 приведены униформ-ротатабельные планы для двух и трех факторов в нормализованных обозначениях.

Каждый фактор в УРП варьируется на пяти уровнях: $-\alpha, -1, 0, +1, +\alpha$ в нормализованных обозначениях. Каждый нормализованный фактор в УРП x_i варьируется в диапазоне от $-\alpha$ до $+\alpha$. Диапазон варьирования натурального фактора X_i : $X_{i \min} \leq X_i \leq X_{i \max}$. Уровню $-\alpha$ соответствует минимальное значение фактора X_i : $X_i^{(-\alpha)} = X_{i \min}$; уровню $+\alpha$ — максимальное значение фактора X_i : $X_i^{(+\alpha)} = X_{i \max}$; $i = 1, 2, \dots, k$; уровню 0 — середина диапазона варьирования:

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

Уровню ($x_1 = -1$) соответствует натуральное значение

$$X_i^{(-1)} = X_i^{(0)} - \frac{X_i^{(+\alpha)} - X_i^{(0)}}{\alpha}.$$

Уровню ($x_1 = +1$) $X_i^{(+1)} = X_i^{(0)} + \frac{X_i^{(+\alpha)} - X_i^{(0)}}{\alpha}.$

Таблица 4.8

Униформ-ротатабельный план для двух факторов

№ опыта	x_1	x_2	y	№ опыта	x_1	x_2	y
Ортогональная часть плана				Звездные точки			
1	+1	-1	y_1	7	-1,414	0	y_7
2	-1	-1	y_2	8	+1,414	0	y_8
3	+1	+1	y_3	Опыты в центре плана			
4	-1	+1	y_4	9	0	0	y_9
Звездные точки				10	0	0	y_{10}
5	-1,414	0	y_5	11	0	0	y_{11}
6	+1,414	0	y_6	12	0	0	y_{12}
				13	0	0	y_{13}

Таблица 4.9

Униформ-ротатабельный план трех факторов

№ опыта	x_1	x_2	x_3	y	№ опыта	x_1	x_2	x_3	y
1	-1	-1	-1	y_1	10	+1,682	0	0	y_{10}
2	+1	-1	-1	y_2	11	0	-1,682	0	y_{11}
3	-1	+1	-1	y_3	12	0	+1,682	0	y_{12}
4	+1	+1	-1	y_4	13	0	0	-1,682	y_{13}
5	-1	-1	+1	y_5	14	0	0	+1,682	y_{14}
6	+1	-1	+1	y_6	15	0	0	0	y_{15}
7	-1	+1	+1	y_7
8	+1	+1	+1	y_8	20	0	0	0	y_{20}
9	-1,682	0	0	y_9					

Переход от натуральных значений факторов к нормализованным осуществляется по тем же формулам, что для ПФП:

$$x_i = \frac{X_i - X_i^{(0)}}{\Delta_i},$$

где Δ_i равно

$$\Delta_i = X_i^{(+1)} - X_i^{(0)} = X_i^{(0)} - X_i^{(-1)}, i = 1, 2, \dots, k.$$

Величина Δ_i не равна половине диапазона варьирования фактора.

Для расчета оценок коэффициентов регрессии равномерн-ротатбельного плана можно воспользоваться формулами, что и для *B*-плана. Коэффициенты T_i для планов с числом факторов $k = 2-5$ можно взять из табл. 4.10.

Таблица 4.10

Коэффициенты T_i для равномерн-ротатбельных планов с числом факторов $k = 2-5$

T_i	Число факторов			
	2	3	4	5 (план с полурепликой)
T_1	0,2	0,16634	0,14285	0,15909
T_2	0,1	0,05679	0,03571	0,03409
T_3	0,125	0,07322	0,04167	0,04167
T_4	0,125	0,04247	0,03125	0,03125
T_5	0,01874	0,0069	0,00372	0,00284
T_6	0,25	0,125	0,0625	0,0625

Опыты, составляющие ортогональную часть, и звездные точки ротатбельного плана при $k \geq 3$ располагаются вблизи сферы в факторном пространстве. Областью варьирования факторов для ротатбельного плана является шар в факторном пространстве (в отличие от куба, например, для *B*-плана). Из-за указанного расположения опытов плана регрессионная модель, построенная по его результатам для $k = 2$, дает удовлетворительное описание внутри круга радиусом α , а в более общем случае при $k \geq 3$ — внутри шара размерности k .

4.4. Исследование регрессионных моделей второго порядка для решения задач оптимизации

Регрессионная модель второго порядка дает весьма полную информацию о влиянии варьируемых факторов на выходную величину. Эту модель можно использовать и для оптимизации условий функционирования объекта, т.е. нахождения значений варьируемых факторов внутри диапазона их варьирования, для которых значение выходной величины оказывается максимальным или минимальным.

Одним из наиболее эффективных методов нахождения оптимальных условий функционирования объекта по уравнению регрессии второго порядка является диссоциативно-шаговый метод [8].

Рассмотрим этот метод. Для однофакторных моделей все возможные варианты решения поставленной задачи описаны в табл. 2.26 [8]. Этот же подход используется при анализе многофакторной модели с диапазонами варьирования -1 — $+1$. Самым простым является случай отсутствия парных взаимодействий: положение точки оптимума по каждому фактору сводится к однофакторному случаю. Для этого последовательно анализируются члены модели, в которые входит сначала только первый фактор, затем только второй, и т.д.

Рассмотрим, например, уравнение (4.8):

$$R_{\text{сж. н.у}} = 46,02 + 17,76(x_1) - 7,74(x_2) + 5,08(x_3) - 3,47(x_1)^2 + 4,43(x_3)^2.$$

Найдем значения факторов, соответствующие максимальной величине отклика, т.е. прочности на сжатие.

Разложение модели на однофакторные модели имеет следующий вид:

$$R_1 = 17,76x_1 - 3,47x_1^2;$$

$$R_2 = -7,74x_2;$$

$$R_3 = 5,08x_3 + 4,43x_3^2.$$

Оптимальные решения по каждой модели определяются по табл. 4.11.

Оптимальные решения для модели

При поиске максимума		Значение $x_{i\text{опт}}$		При поиске минимума	
b_{ii}	b_i	если $ b_i < 2 b_{ii} $ ($x_{i\text{в}}$ внутри диапазона варьирования)	если $ b_i > 2 b_{ii} $ ($x_{i\text{в}}$ вне диапазона варьирования)	b_i	b_{ii}
$b_{ii} < 0$	$b_i < 0$	$x_{i\text{опт}} = x_{i\text{в}} = -b_i/2b_{ii}$	$x_{i\text{опт}} = -1$	$b_i > 0$	$b_{ii} > 0$
	$b_i = 0$		—	$b_i = 0$	
	$b_i > 0$		$x_{i\text{опт}} = +1$	$b_i < 0$	
$b_{ii} = 0$	$b_i < 0$	—	$x_{i\text{опт}} = -1$	$b_i > 0$	$b_{ii} = 0$
	$b_i = 0$	—	—	$b_i = 0$	
	$b_i > 0$	—	$x_{i\text{опт}} = +1$	$b_i < 0$	
$b_{ii} > 0$	$b_i < 0$	$x_{i\text{опт}} = -1$	$x_{i\text{опт}} = -1$	$b_i > 0$	$b_{ii} < 0$
	$b_i = 0$	+1 или -1 (два конкурирующих решения)	—	$b_i = 0$	
	$b_i < 0$	$x_{i\text{опт}} = +1$	$x_{i\text{опт}} = +1$	$b_i < 0$	

Для первой модели: $b_1 = 17,6 > 0$. Так же для этой модели выполняется соотношение $|b_i| \geq \sum_j |b_{ij}|$. Следовательно, согласно табл. 4.11 $x_{1\text{опт}} = +1$. Для второй модели получаем $x_{2\text{опт}} = x_{i\text{опт}} = x_{i\text{в}} = -1$. Для третьей модели — $x_{3\text{опт}} = +1$.

Подставив найденные значения в уравнение (4.8), определим оптимальное значение прочности:

$$R_{\text{сж. н.у}} = 46,02 + 17,76 + 7,74 + 5,08 - 3,47 + 4,43 = 73,6 \text{ МПа.}$$

Проведем анализ многофакторной модели, содержащей k факторов при наличии парных взаимодействий. Согласно диссоциативно-шаговому методу, на основе исходной модели записывается k квазиоднофакторных моделей, т.е. моделей, каждая из которой содержит линейные и квадратичные члены только одного фактора исходной модели, а также парные взаимодействия этого фактора с остальными факторами. Дальнейшее пошаговое изложение относится к случаю отыскания максимума функции отклика. Для нахождения минимума нужно изменить знаки всех коэффициентов регрессии в исходной модели на противоположные.

Шаг 1. Из системы квазиоднофакторных моделей, полученных на основе исследуемой модели, выбирают те, для которых $b_{ii} \geq 0$. Для таких моделей максимум достигается на границе области ва-

рырования фактора x_i , т.е. при $x_i = +1$ или при $x_i = -1$. Если при этом выполняется условие

$$|b_i| \geq \sum_j |b_{ij}|, \quad (4.14)$$

то знак $x_{i\text{опт}}$ совпадает со знаком b_i , а именно:

- если $b_i > 0$, то $x_{i\text{опт}} = +1$;
- если $b_i < 0$, то $x_{i\text{опт}} = -1$.

Шаг 2. Если при выполнении условия $b_{ii} \geq 0$ условие (4.14) не выполняется, то следует рассмотреть две конкурирующие модели: одну, полученную из исходной модели при $x_i = -1$ (с учетом найденных ранее оптимальных факторов), другую — при $x_i = +1$. Дальнейшему исследованию подвергается из двух конкурирующих моделей та модель, для которой сумма абсолютных величин свободного члена и всех линейных коэффициентов регрессии наибольшая, т.е.

$$|b_0| + \sum_i |b_i| \rightarrow \max.$$

Именно по этой модели будет получено максимальное значение отклика.

Шаг 3. Для квазифакторных моделей

$$y_i = b_0 + b_{ii}x_i^2 + b_i x_i + \sum_{i \neq j} b_{ij} x_i x_j, \quad (4.15)$$

у которых $b_{ii} < 0$ и, кроме того, выполняется соотношение

$$|b_i| + \sum_{i \neq j} |b_{ij}| \leq 2|b_{ii}|, \quad (4.16)$$

величина $x_{i\text{опт}}$ определяется из условия равенства нулю производной функции (4.15) по x_i :

$$\frac{\partial y_i}{\partial x_i} = 2b_{ii}x_i + b_i + \sum_{i \neq j} b_{ij}x_j = 0.$$

Отсюда

$$x_{i\text{в}} = x_{i\text{опт}} = \frac{-b_i}{2b_{ii}} - \sum_{i \neq j} \frac{b_{ij}x_j}{2b_{ii}}, \quad (4.17)$$

т.е. $x_{i\text{опт}}$ представляет собой линейную функцию от x_j . Подстановка этой функции в исходную модель позволяет уменьшить на единицу число переменных и перейти к следующему шагу оптимизации.

Шаг 4. Если для модели с отрицательными квадратичными коэффициентами ($b_{ii} < 0$) соотношение (4.16) не выполняется, т.е. имеет место неравенство

$$|b_{ii}| + \sum_{i \neq j} |b_{ij}| > 2|b_{ii}|, \quad (4.18)$$

то проверяют дополнительное условие

$$|b_{ii}| - \sum_{i \neq j} |b_{ij}| \geq 2|b_{ii}|. \quad (4.19)$$

Если оно выполняется, то $x_{i\text{опт}}$ равно или -1 (при $b_i < 0$), или $+1$ (при $b_i > 0$).

Если при выполнении условия (4.18) соотношение (4.19) не имеет места, то для одной части диапазона варьирования фактора x_i величина $x_{i\text{опт}}$ определяется из формулы (4.17), а для другой части $x_{i\text{опт}}$ принимает граничное значение -1 или $+1$.

4.5. Композиционный план

4.5.1. Виды композиционных планов

Композиционные планы используются, как правило, на заключительном этапе исследования, когда модель подбирается последовательно, начиная с простейшего линейного уравнения, которое потом дорабатывается до полной квадратичной формулы. Композиционные планы дают выигрыш по числу опытов по сравнению с другими планами. Эти планы можно применять и при непосредственном построении функции отклика в виде полинома (2.3).

Решение подобных задач основано на применении ортогональных или ротатабельных центральных композиционных планов (ЦКП). Данные планы используют в качестве ядра полный факторный эксперимент или минимально возможные регулярные дробные реплики типа $2k - p$. В качестве дробной реплики применяют такую, в которой два любых парных взаимодействия по модулю не равны друг другу:

$$|x_i x_j| \neq |x_i x_m| \quad (4.20)$$

для любых попарно различных индексов. Именно план ПФП или дробные реплики, удовлетворяющие указанному условию, служат ядром ЦКП.

Широкое практическое применение получили два типа ЦКП, известные как планы Бокса и Хартли. Понятие «центральный» означает, что факторы принимают значения, симметричные относительно центра плана.

Центральный композиционный план второго порядка называют планом Бокса, если его ядром является ПФП 2^k или регулярная реплика типа 2^{k-p} , для которой парные взаимодействия не равны по модулю линейным факторам:

$$x_i \neq \pm x_l x_m, l \neq m; i, l, m = 1, 2, \dots, k$$

и выполняется условие (4.20). Ядром плана Бокса при $k < 5$ является ПФП, а при $k = 5$ может быть ДФП.

План Бокса можно сделать ортогональным либо ротатабельным. Но нельзя добиться одновременного и строго соблюдения обоих свойств.

Центральный композиционный план второго порядка называют планом Хартли, если его ядром является регулярная реплика типа 2^{k-p} , в которой некоторые парные взаимодействия равны по модулю линейным факторам.

Иначе говоря, ЦКП второго порядка будет или планом Бокса или планом Хартли. Планы Хартли обычно более экономны по числу опытов, чем планы Бокса, но уступают им по точности оценивания коэффициентов; кроме того, их нельзя сделать ни ортогональными, ни ротатабельными.

Планы Хартли целесообразно применять, если известно, что часть эффектов b_j или b_{ju} в модели отсутствует (следовательно, простые эффекты можно смешивать с парными взаимодействиями, не теряя в разрешающей способности плана) или тогда, когда дисперсия наблюдений относительно мала.

4.5.2. Ортогональный композиционный план

Построение ортогонального композиционного плана

В планах Бокса к ядру, построенному на основе ПФП или ДФП, добавляется одна точка в центре плана с координатами $(0, 0, \dots, 0)$ и $2k$ «звездных» точек с координатами $(\pm g, 0, \dots, 0), \dots, (0, 0, \dots, \pm g)$.

Построенный таким образом план будет ЦКП второго порядка [10]. Общее количество точек плана при использовании композиционного планирования составит $N = N_0 + 2k + 1$, где N_0 — количество точек ядра плана. В табл. 4.12 и 4.13 содержится описание

соответствующих матриц планирования для ЦКП при $k = 2$. Количество опытов для данного плана $N = 2^2 + 2 \cdot 2 + 1 = 9$. Аналогично строятся ЦКП для произвольного числа факторов, при этом каждый фактор варьируется на пяти уровнях: $-g$; -1 ; 0 ; $1g$.

Таблица 4.12

Матрица ядра плана

Ядро плана	
x_1	x_2
+	+
-	+
+	-
-	-

Таблица 4.13

Матрица дополнительных точек

Дополнительные точки	
x_1	x_2
g	0
$-g$	0
0	g
0	$-g$
0	0

В матрице плана второго порядка не у всех столбцов соблюдается условие симметрии и не все пары столбцов ортогональны. Например, рассмотрим ЦКП второго порядка для трех переменных (табл. 4.14).

Таблица 4.14

ЦКП второго порядка

План	x_0	x_1	x_2	x_3	x_1x_2	x_1x_3	x_2x_3	x_1^2	x_2^2	x_3^2
ПФЭ	+	-	-	-	+	+	+	+	+	+
	+	+	-	-	-	-	+	+	+	+
	+	-	+	-	-	+	-	+	+	+
	+	+	+	-	+	-	-	+	+	+
	+	-	-	+	+	-	-	+	+	+
	+	+	-	+	-	+	-	+	+	+
	+	-	+	+	-	-	+	+	+	+
	+	+	+	+	+	+	+	+	+	+
Звездный план	+	$-g$	0	0	0	0	0	g^2	0	0
	+	g	0	0	0	0	0	g^2	0	0
	+	0	$-g$	0	0	0	0	0	g^2	0
	+	0	g	0	0	0	0	0	g^2	0
	+	0	0	$-g$	0	0	0	0	0	g^2
	+	0	0	g	0	0	0	0	0	g^2
Центр плана	+	0	0	0	0	0	0	0	0	0

Для всех строк плана суммы

$$\sum_{u=1}^N x_{iu}^2 \neq 0, \sum_{u=1}^N x_{iu}^2 x_{ju}^2 \neq 0, \text{ так как } x_{iu}^2 \neq 0.$$

Для устранения асимметрии и нарушений ортогональности ЦКП Бокса необходимо провести преобразование квадратичных параметров и специальным образом выбрать величину плеча g .

Чтобы добиться соблюдения свойства симметричности, следует перейти от x_i^2 к центрированным величинам $x_i^* = x_i^2 - x_{i\text{cp}}^2$ (сумма центрированных величин равна нулю). Среднее значение $x_{i\text{cp}}^2$, как видно из табл. 4.14, для всех x_i^2 одинаково и равно

$$\alpha = \frac{N_0 + 2g^2}{N}. \quad (4.21)$$

Тогда исходную квадратичную модель (2.3) можно преобразовать

$$\begin{aligned} y' &= b_0 + b_1 x_1 + \dots + b_1 x_k + b_{12} x_1 x_2 + \dots + b_{k-1, k} x_{k-1} x_k + \\ &+ b_{11} (x_1^2 - x_{1\text{cp}}^2 + x_{1\text{cp}}^2) + \dots + b_{kk} (x_k^2 - x_{k\text{cp}}^2 + x_{k\text{cp}}^2) = \\ &= d_0 + b_1 x_1 + \dots + b_1 x_k + b_{12} x_1 x_2 + \dots + b_{k-1, k} x_{k-1} x_k + \\ &+ b_{11} x_1^* + \dots + b_{kk} x_k^*, \end{aligned}$$

где $d_0 = b_0 + b_{11} x_{1\text{cp}}^2 + \dots + b_{k-1, k} x_{k\text{cp}}^2 = b_0 + \alpha(b_{11} + \dots + b_{k-1, k})$.

Исходная и преобразованная модели эквивалентны, в них все коэффициенты, за исключением нулевого, совпадают.

После преобразования получим матрицу планирования (табл. 4.15).

Таблица 4.15

Матрица планирования композиционных планов

План	x_0	x_1	x_2	x_3	$x_1 x_2$	$x_1 x_3$	$x_2 x_3$	x_1^*	x_2^*	x_3^*
ПФЭ	+	-	-	-	+	+	+	$1-\alpha$	$1-\alpha$	$1-\alpha$
	+	+	-	-	-	-	+	$1-\alpha$	$1-\alpha$	$1-\alpha$
	+	-	+	-	-	+	-	$1-\alpha$	$1-\alpha$	$1-\alpha$
	+	+	+	-	+	-	-	$1-\alpha$	$1-\alpha$	$1-\alpha$
	+	-	-	+	+	-	-	$1-\alpha$	$1-\alpha$	$1-\alpha$
	+	+	-	+	-	+	-	$1-\alpha$	$1-\alpha$	$1-\alpha$
	+	-	+	+	-	-	+	$1-\alpha$	$1-\alpha$	$1-\alpha$
	+	+	+	+	+	+	+	$1-\alpha$	$1-\alpha$	$1-\alpha$

План	x_0	x_1	x_2	x_3	x_1x_2	x_1x_3	x_2x_3	x_1^*	x_2^*	x_3^*
Звездный план	+	$-g$	0	0	0	0	0	$g^2-\alpha$	$-\alpha$	$-\alpha$
	+	g	0	0	0	0	0	$g^2-\alpha$	$-\alpha$	$-\alpha$
	+	0	$-g$	0	0	0	0	$-\alpha$	$g^2-\alpha$	$-\alpha$
	+	0	g	0	0	0	0	$-\alpha$	$g^2-\alpha$	$-\alpha$
	+	0	0	$-g$	0	0	0	$-\alpha$	$-\alpha$	$g^2-\alpha$
	+	0	0	g	0	0	0	$-\alpha$	$-\alpha$	$g^2-\alpha$
Центр плана	+	0	0	0	0	0	0	$-\alpha$	$-\alpha$	$-\alpha$

Суммы элементов по всем столбцам, за исключением столбца x_0 , равны нулю, т.е. в преобразованной таблице соблюдается свойство симметричности. Столбцы квадратичных членов при произвольных значениях g не являются ортогональными:

$$\sum_{u=1}^N (x_{iu}^2 - \alpha)(x_{ju}^2 - \alpha) = \sum_{u=1}^N x_{iu}^* x_{ju}^* \neq 0, \quad i \neq j.$$

Ортогонализация столбцов

$$\sum_{u=1}^N x_{iu}^* x_{ju}^*,$$

т.е. приравнивание к нулю, достигается специальным выбором величины g . Это значение величины g находится из уравнения

$$\sum_{u=1}^N x_{iu}^* x_{ju}^* = N_0(1-\alpha)^2 - 4\alpha(\gamma^2 - \alpha) + (2k-4)\alpha^2 + \alpha^2 = 0,$$

или

$$\begin{aligned} N_0 - 2\alpha N_0 + N_0 c^2 - 4\alpha g^2 + 4\alpha^2 + 2\alpha^2 - 4\alpha^2 + \alpha^2 = \\ = N_0 - 2(N_0 + 2g^2)\alpha + \alpha^2(N_0 + 2k + 1) = N_0 - 2\alpha^2 N + \alpha^2 N = 0 \end{aligned}$$

Следовательно, $\alpha^2 N = N_0$.

Тогда $\alpha = \sqrt{N_0 / N}$. Подставим найденное значение величины α в уравнение (4.21):

$$\sqrt{\frac{N_0}{N}} = \sqrt{\frac{N_0 + 2g^2}{N}}.$$

Решив уравнение, найдем величину g , которая придает матрице планирования (в том числе табл. 4.15) свойство ортогональности:

$$g = \sqrt{\sqrt{NN_0} - N_0} / 2.$$

Значения g , обеспечивающие ортогональность, например для ядер $2^2, 2^3, 2^4, 2^{5-1}$, составляют соответственно 1; 1,215; 1,414; 1,547.

Оценки коэффициентов регрессии определяются по модифицированной матрице независимых переменных (см. табл. 4.14). В результате ортогональности матрицы все коэффициенты определяются независимо друг от друга по формуле

$$b_i = \frac{\sum_{j=1}^N x_{ij} y_j}{\sum_{j=1}^N x_{ij}^2}.$$

Дисперсии коэффициентов

$$s^2\{b_i\} = \frac{s^2\{y\}}{\sum_{j=1}^N x_{ij}^2},$$

где $s^2\{y\}$ — дисперсия воспроизводимости.

Оценки коэффициентов регрессии определяются по следующим формулам:

$$d_0 = \frac{\sum_{u=1}^N \bar{y}}{N};$$

$$b_0 = d_0 - \alpha \sum_{j=1}^k b_{jj}.$$

Оценки дисперсии коэффициентов регрессии осуществляются по следующим формулам:

$$\sigma^2(b_i) = \sum_{u=1}^N x_{iu}^2 \sigma^2(\bar{y}_u) / \left[\sum_{u=1}^N x_{iu}^2 \right];$$

$$\sigma^2(d_0) = \sum_{u=1}^N \sigma^2(\bar{y}_u) / N^2;$$

$$\sigma^2(b_i) = \sigma^2(d_0) + \alpha^2 \sum_{j=1}^k \sigma^2(b_{ij}),$$

где $\sigma^2(\bar{y}_u)$ — оценка дисперсии среднего значения функции отклика в u -й точке плана.

Оценки дисперсии коэффициентов являются различными, так как вычисляются по разным совокупностям точек плана. Оценка дисперсии функции отклика зависит не только от расстояния до заданной точки от центра, но и от ее положения в пространстве, т.е. ортогональный план второго порядка не является ротатабельным.

Проверка однородности дисперсии воспроизводимости, адекватности модели, значимости коэффициентов полинома в случае применения ортогональных ЦКП второго порядка осуществляется по рассмотренной ранее схеме (п. 2.3.3).

Проверка адекватности экспериментальной модели

Значимость каждого коэффициента по критерию Стьюдента определяется по формуле

$$t_{\text{кр}} = \frac{|b_i|}{s_i}.$$

Коэффициент b_i считается значимым, если $t_{\text{кр}} > t_p(f_2)$.

Оценки дисперсии коэффициентов s_i определяются следующим образом:

$$s_i^2 = \begin{cases} s^2 b_0 = \frac{s^2}{N}, \\ s^2 b_i = \frac{s^2}{2^n + 2\alpha^2}, \\ s^2 b_{ij} = \frac{s^2}{2^n}, \\ s^2 b_{jj} = \frac{s^2}{2^n(1-g)^2 + 2(\alpha^2 - g)^2 + (2n-1)g^2}, \end{cases}$$

где s^2 — оценка дисперсии ошибок наблюдения;

$$s^2 = \frac{\sum_{i=1}^{n_0} (y_i^0 - \bar{y}^0)^2}{n_0 - 1},$$

здесь n_0 — число повторяющихся экспериментов в центре;

y_i^0 — полученное значение i -го эксперимента в центре;

\bar{y}^0 — среднее значение n_0 экспериментов в центре.

Адекватность экспериментальной модели проверяется по критерию Фишера по формуле

$$F_{\text{расч}} = \frac{s_R^2}{s^2},$$

где $s_R^2 = \frac{1}{N - (k + 1)} \sum_{i=1}^N (y_i - \hat{y}_i)^2,$

здесь k — число коэффициентов, обладающих значимостью;

y_i — наблюдаемое значение i -го эксперимента;

\hat{y}_i — полученное значение из экспериментальной функции в соответствии с i -м экспериментом.

Условием адекватности модели является

$$F_{\text{расч}} \leq F_{\text{кр}}(f_1/f_2),$$

где f_1 — число степеней свободы для остаточной дисперсии,
 $f_1 = N - (k + 1);$

f_2 — число степеней свободы для оценки дисперсии наблюдений,
 $f_2 = n_0 - 1.$

Если $F_{\text{расч}} \leq F_{\text{кр}}(f_1 f_2)$, то модель адекватна; если $F_{\text{расч}} \geq F_{\text{кр}}$, то гипотеза об адекватности модели отклоняется. Значение $F_{\text{кр}}(f_1 f_2)$ берется из табл. П4 приложения.

4.6. Применение композиционного плана

Одним из возможных вариантов подбора состава мелкозернистого гидротехнического бетона с добавками является метод математического планирования экспериментов по ортогональному центральному плану второго порядка.

С целью повышения водонепроницаемости и морозостойкости гидротехнического мелкозернистого бетона в бетонную смесь могут вводиться гидрофобизирующие кремнийорганические жидкости [3]: «СОФЭКСИЛ-40», представляющая собой 50%-й водный концентрат метилсиликоната калия, и «СОФЭКСИЛ 60-80» в виде 60%-й водной силан-силоксановой эмульсии.

Пример 4.2¹. В результате проведенной экспериментальной работы было установлено, что наибольшая прочность мелкозернистого бетона на сжатие (55,4 МПа) и на растяжение при изгибе (8,8 МПа) в возрасте 28 суток нормального твердения, а также его высокая водонепроницаемость (W20) достигаются при использовании суперпластификатора «Химком Ф-1» в количестве 1,2 % от массы цемента в пересчете на сухое вещество (табл. 4.16).

Таблица 4.16

Физико-механические показатели мелкозернистого бетона, пластифицированного «Химком Ф-1»

Состав цемента и его показатели	№ состава			
	I	II	III	IV
Цемент, кг	550	550	550	550
Песок, кг	1500	1500	1500	1500
Химком Ф-1 (30%-й раствор), кг	11(0,6 %)	16(0,9 %)	22(1,2 %)	28(1,5 %)
Вода, кг	180	156	150	144
Подвижность по расплыву конуса, мм	160	115	150	159

¹ Пример составлен на основе эксперимента из диссертационной работы С.В. Алексашина [9].

Состав цемента и его показатели	№ состава			
	I	II	III	IV
$R_{изг28}$, МПа	5,13	7,5	8,8	8,3
$R_{сж28}$, МПа	26,4	46,2	55,4	53,2
Водонепроницаемость, с/см ³ / марка	—	—	355,0/W20	157,9/W20

Примечание. В скобках указана концентрация «Химком Ф-1» в % от массы цемента в пересчете на сухое вещество.

Требуется определить зависимость водонепроницаемости, прочности при сжатии и растяжении и фактической морозостойкости мелкозернистого бетона от расхода метаксаолина и кремнийорганического модификатора «СОФЭКСИЛ-40».

Решение

Установлено, что введение в пластифицированную мелкозернистую бетонную смесь «СОФЭКСИЛ-40» в количестве 0,2 % от массы цемента эффективнее повышает морозостойкость и водонепроницаемость бетона, чем добавка 1,5 % «СОФЭКСИЛ 60-80» (табл. 4.17).

Таблица 4.17

Определение прочности, морозостойкости и водонепроницаемости пластифицированного мелкозернистого бетона, гидрофобизированного кремнийорганическими жидкостями

Состав цемента и его показатели	№ состава			
	I (конт- рольный)	II	III	IV
Цемент, кг	550	550	550	550
Песок, кг	1500	1417,5	1417,5	1417,5
«Химком Ф-1» (30%-й раствор), кг	22(1,2 %)	22(1,2 %)	22(1,2 %)	22(1,2 %)
Вода, кг	150	150	150	150
Метаксаолин, кг	—	82,5(15 %)	82,5(15 %)	82,5(15 %)
«СОФЭКСИЛ-40», кг	—	—	1,1(0,2%)	—
«Софэксил 60-80», кг	—	—	—	8,25(1,5 %)

Состав цемента и его показатели	№ состава			
	I (конт- рольный)	II	III	IV
$R_{сж28}$, МПа	55,4	56,7	57,5	51,4
Фактическая морозостойкость, циклы испытаний	492	446	653	379
Марка по морозостойкости	F400	F400	F600	F300
Водонепроницаемость, с/см ³ / марка	355,0/W20	324,8/W20	338,5/W20	296,7/W20

Примечание. В скобках указана концентрация добавок в % от массы цемента.

Для дальнейших исследований принимаем расход пластифицирующей добавки постоянным и равным 1,2 % от массы вяжущего и В/Ц равное 0,31 и будем использовать гидрофобизирующую добавку «СОФЭКСИЛ-40».

К мелкозернистому гидротехническому бетону проектируемого состава предъявляются следующие требования:

- прочность при сжатии 50 МПа;
- водонепроницаемость W20;
- повышенная морозостойкость.

Подбор оптимального состава мелкозернистого гидротехнического бетона производим путем ортогонального центрального планирования второго порядка.

Факторы, влияющие на прочность, водонепроницаемость и морозостойкость мелкозернистого бетона, приведены в табл. 4.18.

Таблица 4.18

Уровни факторов и интервалы варьирования

Факторы		Уровни варьирования			Интервал варьирования
В натуральном виде	В нормализованном виде	-1	0	+1	
Расход метаксаолина	x_1	5	10	15	5
Расход «СОФЭКСИЛ-40»	x_2	0,08	0,14	0,21	0,06

В табл. 4.19 приведены экспериментальные значения следующих параметров оптимизации:

y_1 — водонепроницаемости M (по сопротивлению прониканию воздуха);

y_2 — прочности при сжатии $R_{сж28}$;

y_3 — прочности при растяжении на изгиб $R_{изг28}$;

y_4 — фактической морозостойкости F .

Результаты расчетов коэффициентов регрессии представлены в табл. 4.20.

По результатам расчетов получены следующие уравнения:

$$y_1 = 215 + 68,33x_1 + 22x_2 + 86,09(x_1^2 - 0,666) - 5,01(x_2^2 - 0,666) + 8,25x_1x_2;$$

$$y_2 = 53,92 + 2,25x_1 + 0,03x_2 + 0,98(x_1^2 - 0,666) - 0,67(x_2^2 - 0,666) + 0,98x_1x_2;$$

$$y_3 = 7,76 + 0,77x_1 + 0,07x_2 + 0,13(x_1^2 - 0,666) + 0,23(x_2^2 - 0,666) + 0,15x_1x_2;$$

$$y_4 = 436,86 + 102,17x_1 + 40,83x_2 + 73,91(x_1^2 - 0,666) - 3,17(x_2^2 - 0,666) + 44x_1x_2.$$

После проверки значимости коэффициентов незначимые коэффициенты были отброшены и получены уравнения следующего вида:

$$y_1 = 161 + 68x_1 + 22x_2 + 86x_1^2 - 5x_2^2 + 8,25x_1x_2;$$

$$y_2 = 54 + 2,25x_1 + x_1^2 - 0,5x_2^2 + x_1x_2;$$

$$y_3 = 7,6 + 0,8x_1 + 0,1x_2 + 0,1x_1^2 + 0,2x_2^2 + 0,1x_1x_2;$$

$$y_4 = 390 + 102x_1 + 40x_2 + 74x_1^2 - 3x_2^2 + 44x_1x_2.$$

Для проверки адекватности полученных выше уравнений регрессии используются табл. 4.21—4.24. В табл. 4.21 приведены значения для расчета адекватности модели зависимости водонепроницаемости.

В результате проверки определено, что $F_{расч} = 7,17$. Для $f_1 = 4$ и $f_2 = 3$ критическое значение F -критерия $F_{0,05} = 9,12$. Так как $F_{расч} < F_{0,05}$, следовательно, зависимость y_1 удовлетворяет результатам практического эксперимента.

Ортогональный план второго порядка и результаты экспериментов

№ п/п	Значения факторов					Параметры оптимизации									
	нормализованные		натуральные			$M, \text{с/см}^3$	$R_{\text{жжЗ}}, \text{МПа}$		$R_{\text{плтЗ}}, \text{МПа}$		$F, \text{циклы}$	$M_{\text{ср}}, \text{с/см}^3$	$R_{\text{жжЗ}}, \text{МПа}$	$R_{\text{плтЗ}}, \text{МПа}$	$F, \text{циклы}$
	x_1	x_2	$x_1^2 - g$	$x_2^2 - g$	$x_1 x_2$		$\Gamma\Phi$	y_1	y_2	y_1					
1	+1	+1	+0,333	+0,333	15	0,2	325	56,2	8,2	618					
2	+1	+1	+0,333	+0,333	15	0,2	346	58,5	9,4	686	338	57,5	8,9	653	
3					15	0,2	343	57,8	9,1	655					
4	+1	0	+0,333	-0,666	15	0,14	325	56,3	8,4	583	322	56,8	8,5	578	
5	+1	0	+0,333	-0,666	15	0,14	331	57,2	8,9	574					
6					15	0,14	310	56,9	8,2	577					
7					15	0,08	286	56,4	8,6	479					
8	+1	-1	+0,333	+0,333	15	0,08	280	54,1	8,4	456	276	55,2	8,3	460	
9					15	0,08	262	55,1	7,9	445					
10	0	+1	-0,666	+0,333	10	0,2	178	52,3	7,8	403	173	52,7	7,7	401	
11	0	+1	-0,666	+0,333	10	0,2	156	51,4	7,4	382					
12					10	0,2	185	54,4	7,9	418					
13	0	0	-0,666	-0,666	10	0,14	152	52,8	7,2	386	168	53,9	7,4	396	
14	0	0	-0,666	-0,666	10	0,14	168	54,2	7,6	392					
15					10	0,14	184	54,7	7,4	410					
16					10	0,08	141	53,2	8,1	381	132	53,2	7,9	366	
17	0	-1	-0,666	+0,333	10	0,08	135	54,3	7,7	348					
18					10	0,08	120	52,1	7,9	369					
19					5	0,2	201	51,4	7,1	384	195	51	7,1	376	
20	-1	+1	+0,333	+0,333	5	0,2	186	50,1	6,9	375					
21					5	0,2	198	51,5	7,3	369					
22					5	0,14	168	51,8	6,7	354	165	52,4	6,9	343	
23	-1	0	+0,333	-0,666	5	0,14	152	52,3	7,1	326					
24					5	0,14	175	53,1	6,9	349					
25					5	0,08	158	51,9	6,6	338	166	52,6	7,1	359	
26	-1	-1	+0,333	+0,333	5	0,08	176	52,6	7,5	396					
27					5	0,08	164	53,3	7,2	343					
	0	0			270	3,78	5805	1455,9	209,4	11796	1935	485,3	69,8	3932	

Коэффициенты регрессии

Коэффициент регрессии	Значения коэффициентов регрессии			
	y_1	y_2	y_3	y_4
b_0	215,00	53,92	7,76	436,89
b_1	68,33	2,25	0,77	102,17
b_2	22,00	0,03	0,07	40,83
b_{11}	86,09	0,98	0,13	73,91
b_{22}	-5,01	-0,67	0,23	-3,17
b_{12}	8,25	0,98	0,15	44,00

Таблица 4.21

Фактическая и рассчитанная водонепроницаемость

№ п/п	y_{1i}	\tilde{y}_{1i}	$(y_{1i} - \tilde{y}_{1i})^2$
1	338	340,25	5,06
2	322	323,25	1,56
3	276	279,75	14,06
4	173	178	25,00
5	168	161	49,00
6	132	134	4,00
7	195	187,75	52,56
8	165	179	196,00
9	166	160,25	33,06

В табл. 4.22 приведены значения фактической и рассчитанной прочности бетона на сжатие.

Таблица 4.22

Фактическая и рассчитанная прочность бетона на сжатие

№ п/п	y_{2i}	\tilde{y}_{2i}	$(y_{2i} - \tilde{y}_{2i})^2$
1	57,6	57,74	0,06
2	56,7	57,35	0,20
3	55,3	55,65	0,30
4	52,8	53,6	0,64
5	53,8	54,1	0,01
6	53,3	53,4	0,09
7	51,1	51,15	0,06
8	52,5	52,65	0,12
9	52,65	53,3	0,42

В результате проверки определено, что $F_{\text{расч}} = 1,37$. Для $f_1 = 4$ и $f_2 = 3$ критическое значение F -критерия $F_{0,05} = 9,12$. Так как $F_{\text{расч}} < F_{0,05}$, то зависимость y_2 удовлетворяет результатам практического эксперимента.

В табл. 4.23 представлены значения фактической и рассчитанной прочности бетона на растяжение при изгибе.

Таблица 4.23

Фактическая и рассчитанная прочность бетона на растяжение при изгибе

№ п/п	y_{3i}	\hat{y}_{3i}	$(y_{3i} - \hat{y}_{3i})^2$
1	8,8	8,8	0,00
2	8,6	8,6	0,00
3	8,4	8,6	0,04
4	7,6	7,8	0,04
5	7,5	7,7	0,04
6	7,9	7,7	0,04
7	7,1	7,1	0,00
8	6,8	6,8	0,00
9	7,2	7,2	0,00

В результате проверки определено, что $F_{\text{расч}} = 0,11$. Для $f_1 = 4$ и $f_2 = 3$ критическое значение F -критерия $F_{0,05} = 9,12$. Так как $F_{\text{расч}} < F_{0,05}$, то зависимость y_3 удовлетворяет результатам практического эксперимента.

В табл. 4.24 приведены значения фактической и рассчитанной морозостойкости бетона.

Таблица 4.24

Фактическая и рассчитанная морозостойкость бетона

№ п/п	y_{4i}	\hat{y}_{4i}	$(y_{4i} - \hat{y}_{4i})^2$
1	653	647	36,00
2	578	566	144,00
3	460	479	361,00
4	401	427	676,00
5	396	390	36,00
6	366	347	361,00
7	376	355	441,00
8	343	362	361,00
9	359	363	16,00

В результате проверки определено, что $F_{\text{расч}} = 2,23$. Для $f_1 = 4$ и $f_2 = 3$ критическое значение F -критерия $F_{0,05} = 9,12$. Так как $F_{\text{расч}} < F_{0,05}$, то зависимость y_4 удовлетворяет результатам практического эксперимента.

Исходя из полученных уравнений, описывающих зависимость свойств мелкозернистого бетона от дозировки его компонентов, построены графики (рис. 4.6—4.9).

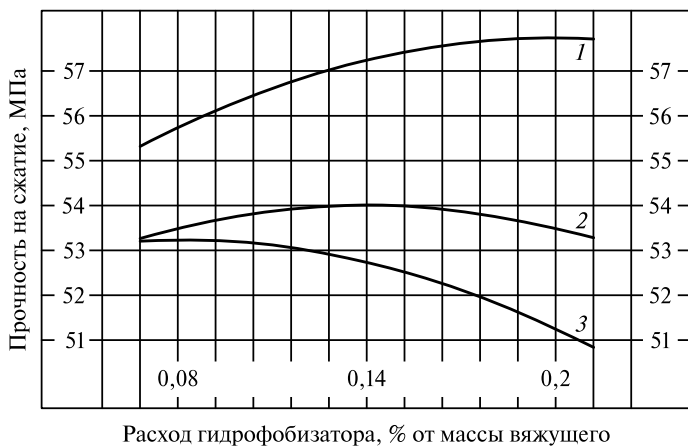


Рис. 4.6. Зависимость прочности при сжатии мелкозернистого бетона от расхода гидрофобизирующей добавки «СОФЭКСИЛ-40» при расходе высокоактивного метаксаолина 5 (1), 10 (2) и 15 (3) %, соответственно

Введение в мелкозернистый бетон высокоактивного метаксаолина оказывает влияние на его прочность при сжатии. Так, из рис. 4.6 видно, что кривая 1, показывающая прочность при сжатии при расходе метаксаолина $82,5 \text{ кг/м}^3$ (15 %), вся лежит выше кривой 2, показывающей прочность при сжатии при расходе метаксаолина 55 кг/м^3 (10 %), которая, в свою очередь, располагается выше кривой 3, показывающей прочность при сжатии при расходе $27,5 \text{ кг/м}^3$ (5 %). Следовательно, прочность при сжатии возрастает с увеличением дозировки метаксаолина, однако его содержание необходимо ограничивать 15 % от массы цемента, так как он содержит активный диоксид кремния и оксид алюминия примерно в одинаковой пропорции и по этой причине сильнее, чем микрокремнезем, связывает свободный гидроксид кальция, что приводит к снижению щелочности среды в бетоне и может вызвать коррозию стальной арматуры.

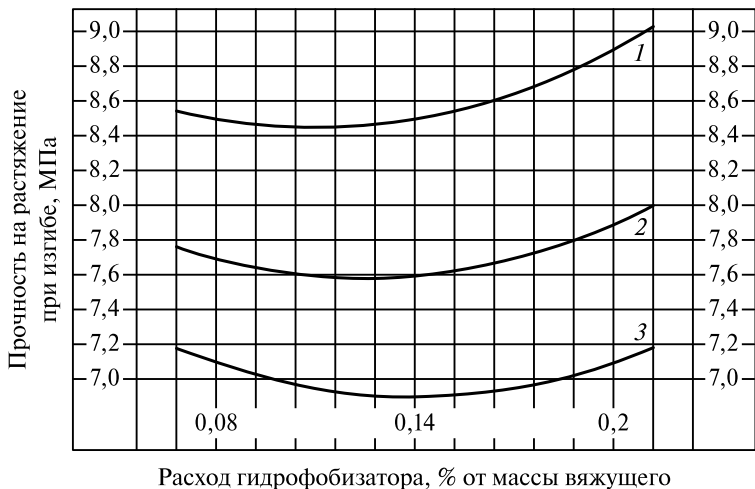


Рис. 4.7. Зависимость прочности на растяжение при изгибе мелкозернистого бетона от расхода гидрофобизирующей добавки «СОФЭКСИЛ-40» при расходе высокоактивного метакаолина 5 (1), 10 (2) и 15 (3) %, соответственно

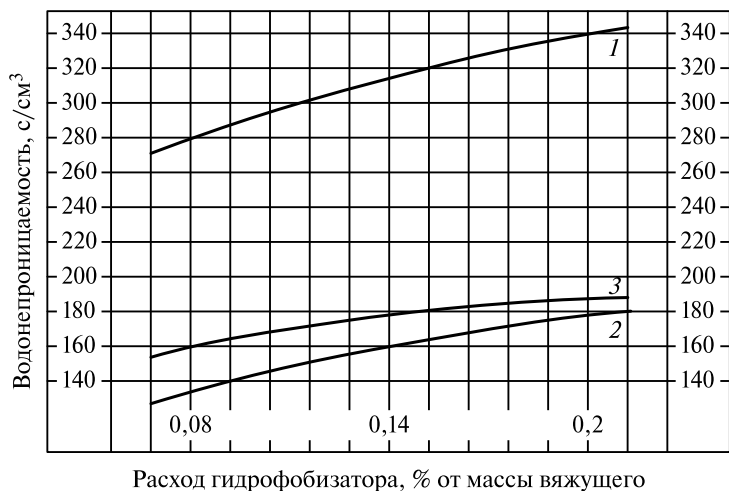


Рис. 4.8. Зависимость водонепроницаемости мелкозернистого бетона от расхода гидрофобизирующей добавки «СОФЭКСИЛ-40» при расходе высокоактивного метакаолина 5 (1), 10 (2) и 15 (3) %, соответственно

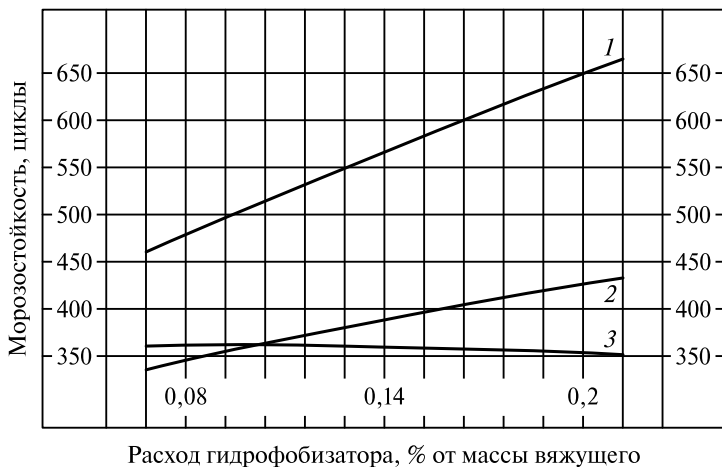


Рис. 4.9. Зависимость морозостойкости мелкозернистого бетона от расхода гидрофобизирующей добавки «СОФЭКСИЛ-40» при расходе высокоактивного метаксаолина 5 (1), 10 (2) и 15 (3) %, соответственно

Также на графиках видно, что при введении гидрофобизирующей добавки «СОФЭКСИЛ-40» мелкозернистый бетон показывает небольшой стабильный прирост прочности только при самой высокой концентрации метаксаолина. Например, при концентрации метаксаолина 15% рост прочности мелкозернистого бетона на сжатие составил 4% при расходе гидрофобизирующей добавки 0,2% от массы вяжущего.

При построении зависимостей прочности на растяжение при изгибе от содержания метаксаолина и гидрофобизатора наблюдаются те же закономерности (см. рис. 4.7).

Одним из самых важных факторов, влияющих на долговечность гидротехнического бетона, является водонепроницаемость. Чем больше водонепроницаемость бетона, тем меньшее воздействие оказывает на него как сама вода, так и растворенные в ней агрессивные вещества.

Метаксаолин является микронаполнителем и имеет высокую удельную площадь поверхности, достигающую 30 м²/г. Введение метаксаолина в мелкозернистый бетон повышает его плотность, а следовательно, и водонепроницаемость. Из графиков рис. 4.8 видно, что при дозировках высокоактивного метаксаолина 5 и 10% от массы вяжущего водонепроницаемость остается примерно на одина-

ковом уровне, однако при увеличении дозировки метакаолина до 15 % наблюдается рост водонепроницаемости практически в два раза. Также видно, что увеличение дозировки гидрофобизирующей добавки с 0,08 до 0,2 % приводит к росту водонепроницаемости в среднем на 20—40% в зависимости от концентрации высокоактивного метакаолина и в точке максимума составляет 340 с/см^3 , что соответствует классу бетона по водонепроницаемости W20 и более.

В климатических условиях Российской Федерации важнейшим показателем долговечности бетона является его морозостойкость. Она зависит от многих факторов, в числе которых прочность бетона, его плотность, гранулометрический состав, пористость, тип и размеры пор, водонепроницаемость. На рис. 4.9 видно, что при увеличении концентрации метакаолина морозостойкость мелкозернистого бетона возрастает. Это связано со значительным уплотнением микроструктуры бетона, а также с ростом его прочности. Также на графиках видно, что увеличение концентрации гидрофобизирующей добавки «СОФЭКСИЛ-40» приводит к росту морозостойкости на 30—40% в зависимости от расхода высокоактивного метакаолина. При дозировке метакаолина 15 % и дозировке гидрофобизатора 0,2 % от массы вяжущего морозостойкость составляет более 650 циклов, что соответствует марке по морозостойкости F600.

Библиографический список

1. *Вентцель Е.С., Овчаров Л.А.* Теория вероятностей и ее инженерные приложения. 4-е изд., стер. М. : Высшая школа, 2007. 491 с.
2. *Колемаев В.А., Староверов О.В., Турундаевский В.Б.* Теория вероятностей и математическая статистика. М. : Высшая школа, 1991. 399 с.
3. Теория вероятностей : курс лекций / Л.В. Кирьянова, А.Ю. Лемин, А.Г. Мясников [и др.]. М. : МГСУ, 2012. 87 с.
4. *Кремер Н.Ш.* Теория вероятностей и математическая статистика. 2-е изд., перераб. и доп. М. : ЮНИТИ, 2004. 573 с.
5. *Ивченко Г.И., Медведев Ю.И.* Математическая статистика : учеб. пособие для втузов. 2-е изд., доп. М. : Высшая школа, 1992. 304 с.
6. *Кирьянова Л.В., Лемин А.Ю., Мясников А.Г.* Математическая статистика : курс лекций. М. : МГСУ, 2014. 116 с.
7. *Пижурич А.А., Розенблит М.С.* Исследования процессов деревообработки. М. : Лесная промышленность, 1984. 232 с.
8. *Соловьев В.Г.* Влияние дисперсного армирования на структурообразование и прочностные свойства сталефибробетонов, подвергаемых тепловлажностной обработке : автореф. дис. ... канд. техн. наук / Воронежский гос. архитектурно-строит. ун-т. Липецк, 2009.
9. *Алексашин С.В., Булгаков Б.И., Попова М.Н.* Повышение эксплуатационных свойств пластифицированных гидротехнических мелкозернистых бетонов. Подбор оптимального состава // Известия ЮФУ. Серия: Технические науки. 2014. № 1. С. 195—201.
10. *Батрак А.П.* Планирование и организация эксперимента : учеб. пособие к изучению теоретического курса для студентов направления 220500. Красноярск : ИПЦ СФУ, 2012. 60 с.

Значения *t*-критерия Стьюдента

<i>n</i>	<i>p</i>														
	0,80	0,90	0,95	0,98	0,99	0,995	0,998	0,999							
1	3,0770	6,3130	12,7060	31,820	63,656	127,656	318,306	636,619							
2	1,8850	2,9200	4,3020	6,964	9,924	14,089	22,327	31,599							
3	1,6377	2,35340	3,182	4,540	5,840	7,458	10,214	12,924							
4	1,5332	2,13180	2,776	3,746	4,604	5,597	7,173	8,610							
5	1,4759	2,01500	2,570	3,649	4,0321	4,773	5,893	6,863							
6	1,4390	1,943	2,4460	3,1420	3,7070	4,316	5,2070	5,958							
7	1,4149	1,8946	2,3646	2,998	3,4995	4,2293	4,785	5,4079							
8	1,3968	1,8596	2,3060	2,8965	3,3554	3,832	4,5008	5,0413							
9	1,3830	1,8331	2,2622	2,8214	3,2498	3,6897	4,2968	4,780							
10	1,3720	1,8125	2,2281	2,7638	3,1693	3,5814	4,1437	4,5869							
11	1,363	1,795	2,201	2,718	3,105	3,496	4,024	4,437							
12	1,3562	1,7823	2,1788	2,6810	3,0845	3,4284	3,929	4,178							
13	1,3502	1,7709	2,1604	2,6503	3,1123	3,3725	3,852	4,220							
14	1,3450	1,7613	2,1448	2,6245	2,976	3,3257	3,787	4,140							
15	1,3406	1,7530	2,1314	2,6025	2,9467	3,2860	3,732	4,072							

16	1,3360	1,7450	2,1190	2,5830	2,9200	3,2520	3,6860	4,0150
17	1,3334	1,7396	2,1098	2,5668	2,8982	3,2224	3,6458	3,965
18	1,3304	1,7341	2,1009	2,5514	2,8784	3,1966	3,6105	3,9216
19	1,3277	1,7291	2,0930	2,5395	2,8609	3,1737	3,5794	3,8834
20	1,3253	1,7247	2,08600	2,5280	2,8453	3,1534	3,5518	3,8495
21	1,3230	1,7200	2,2,0790	2,5170	2,8310	3,1350	3,5270	3,8190
22	1,3212	1,7117	2,0739	2,5083	2,8188	3,1188	3,5050	3,7921
23	1,3195	1,7139	2,0687	2,4999	2,8073	3,1040	3,4850	3,7676
24	1,3178	1,7109	2,0639	2,4922	2,7969	3,0905	3,4668	3,7454
25	1,3163	1,7081	2,0595	2,4851	2,7874	3,0782	3,4502	3,7251
26	1,315	1,705	2,059	2,478	2,778	3,0660	3,4360	3,7060
27	1,3137	1,7033	2,0518	2,4727	2,7707	3,0565	3,4210	3,6896
28	1,3125	1,7011	2,0484	2,4671	2,7633	3,0469	3,4082	3,6739
29	1,3114	1,6991	2,0452	2,4620	2,7564	3,0360	3,3962	3,8494
30	1,3104	1,6973	2,0423	2,4573	2,7500	3,0298	3,3852	3,6460
32	1,3080	1,6930	2,0360	2,4480	2,7380	3,0140	3,3650	3,6210
34	1,3070	1,6909	2,0322	2,4411	2,7284	3,9520	3,3479	3,6007
36	1,3050	1,6883	2,0281	2,4345	2,7195	9,490	3,3326	3,5821
38	1,3042	1,6860	2,0244	2,4286	2,7116	3,9808	3,3190	3,5657
40	1,303	1,6839	2,0211	2,4233	2,7045	3,9712	3,3069	3,5510
42	1,320	1,682	2,018	2,418	2,6980	2,6930	3,2960	3,5370

<i>n</i>	<i>p</i>									
	0,80	0,90	0,95	0,98	0,99	0,995	0,998	0,999		
44	1,301	1,6802	2,0154	2,4141	2,6923	3,9555	3,2861	3,5258		
46	1,300	1,6767	2,0129	2,4102	2,6870	3,9488	3,2771	3,5150		
48	1,299	1,6772	2,0106	2,4056	2,6822	3,9426	3,2689	3,5051		
50	1,298	1,6759	2,0086	2,4033	2,6778	3,9370	3,2614	3,4060		
55	1,2997	1,673	2,0040	2,3960	2,6680	2,9240	3,2560	3,4760		
60	1,2958	1,6706	2,0003	2,3901	2,6603	3,9146	3,2317	3,4602		
65	1,2947	1,6686	1,997	2,3851	2,6536	3,9060	3,2204	3,4466		
70	1,2938	1,6689	1,9944	2,3808	2,6479	3,8987	3,2108	3,4350		
80	1,2820	1,6640	1,9900	2,3730	2,6380	2,8870	3,1950	3,4160		
90	1,2910	1,6620	1,9867	2,3885	2,6316	2,8779	3,1833	3,4019		
100	1,2901	1,6602	1,9840	2,3642	2,6259	2,8707	3,1737	3,3905		
120	1,2888	1,6577	1,9719	2,3578	2,6174	2,8598	3,1595	3,3735		
150	1,2872	1,6551	1,9759	2,3515	2,6090	2,8482	3,1455	3,3566		
200	1,2858	1,6525	1,9719	2,3451	2,6006	2,8385	3,1315	3,3398		
250	1,2849	1,6510	1,9695	2,3414	2,5966	2,8222	3,1232	3,3299		
300	1,2844	1,6499	1,9679	2,3388	2,5923	2,8279	3,1176	3,3233		
400	1,2837	1,6487	1,9659	2,3357	2,5882	2,8227	3,1107	3,3150		
500	1,2830	1,6470	1,9640	2,3330	2,7850	2,8190	3,1060	3,3100		

Значения *t*-критерия Стьюдента

<i>f</i>	<i>q</i>		<i>f</i>	<i>q</i>		<i>f</i>	<i>q</i>	
	0,01	0,05		0,01	0,05		0,01	0,05
1	63,66	12,71	14	2,98	12,71	27	2,77	2,05
2	9,92	4,30	15	2,95	4,30	28	2,76	2,05
3	5,84	3,18	16	2,92	3,18	29	2,76	2,05
4	4,60	2,78	17	2,90	2,78	30	2,75	2,04
5	4,03	2,57	18	2,88	2,57	40	2,70	2,02
6	3,71	2,45	19	2,86	2,45	50	2,68	2,01
7	3,50	2,36	20	2,85	2,36	60	2,66	2,00
8	3,36	2,31	21	2,83	2,31	80	2,64	1,99
9	3,25	2,26	22	2,82	2,26	100	2,63	1,98
10	3,17	2,23	23	2,81	2,23	120	2,62	1,98
11	3,11	2,20	24	2,80	2,20	200	2,60	1,97
12	3,05	2,18	25	2,79	2,18	500	2,59	1,96
13	3,01	2,16	26	2,78	2,16	∞	2,58	1,96

Значения критерия Пирсона

f	Вероятность $P(\chi^2(f) > \chi^2_\gamma) = \gamma$															
	0,99	0,98	0,95	0,90	0,80	0,20	0,10	0,05	0,02	0,01						
5	0,554	0,752	1,145	1,610	2,343	7,289	9,236	11,070	13,388	15,086						
6	0,872	1,134	1,635	2,204	3,070	8,558	10,645	12,592	15,033	16,812						
7	1,239	1,564	2,167	2,833	3,822	9,803	12,017	14,067	16,622	18,475						
8	1,646	2,032	2,733	3,490	4,594	11,030	13,362	15,507	18,168	20,090						
9	2,088	2,532	3,325	4,168	5,380	12,242	14,684	16,919	19,679	21,666						
10	2,558	3,059	3,940	4,865	6,179	13,442	15,987	18,307	21,161	23,209						
11	3,053	3,609	4,575	5,578	6,989	14,631	17,275	19,675	22,618	24,725						
12	3,571	4,178	5,226	6,304	7,807	15,812	18,549	21,026	24,054	26,217						
13	4,107	4,765	5,892	7,042	8,634	16,985	19,812	22,362	25,472	27,688						
14	4,660	5,368	6,571	7,790	9,467	18,151	21,064	23,685	26,873	29,141						
15	5,229	5,985	7,262	8,547	10,307	19,311	22,307	24,996	28,259	30,578						
16	5,812	6,614	7,962	9,312	11,152	20,465	23,542	26,296	29,633	32,000						
17	6,408	7,255	8,672	10,085	12,002	21,615	24,769	27,587	30,995	33,409						

18	7,015	7,906	9,390	10,865	12,857	22,760	25,989	28,869	32,346	34,805
19	7,633	8,567	10,117	11,651	13,716	23,900	27,204	30,144	33,687	36,191
20	8,260	9,237	10,851	12,443	14,578	25,038	28,412	31,410	35,020	37,566
21	8,897	9,915	11,591	13,240	15,445	26,171	29,615	32,671	36,343	38,932
22	9,542	10,600	12,338	14,041	16,314	27,301	30,813	33,924	37,659	40,289
23	10,196	11,293	13,091	14,848	17,187	28,429	32,007	35,172	38,968	41,638
24	10,856	11,992	13,848	15,659	18,062	29,553	33,196	36,415	40,270	42,980
25	11,524	12,697	14,611	16,473	18,940	30,675	34,382	37,652	41,566	44,314
26	12,198	13,409	15,379	17,292	19,820	31,795	35,563	38,885	42,856	45,642
27	12,879	14,125	16,151	18,114	20,703	32,912	36,741	40,113	44,140	46,963
28	13,565	14,847	16,928	18,939	21,588	34,027	37,916	41,337	45,419	48,278
29	14,256	15,574	17,708	19,768	22,475	35,139	39,087	42,557	46,693	49,588
30	14,953	16,306	18,493	20,599	23,364	36,250	40,256	43,773	47,962	50,892

Значения F -критерия Фишера

f_2		f_1														
		1	2	3	4	5	6	7	8	9	10	12	15	20	30	∞
$q = 0,05$																
1	161	200	216	225	230	234	237	239	241	242	244	246	248	250	254	254
2	18,51	19,00	19,16	19,25	19,30	19,33	19,35	19,37	19,38	19,40	19,41	19,43	19,45	19,46	19,50	19,50
3	10,13	9,55	9,28	9,12	9,01	8,94	8,89	8,85	8,81	8,79	8,74	8,70	8,66	8,62	8,53	8,53
4	7,71	6,94	6,59	6,39	6,26	6,16	6,09	6,04	6,00	5,94	5,91	5,86	5,80	5,75	5,63	5,63
5	6,61	5,79	5,41	5,19	5,05	4,95	4,88	4,82	4,77	4,74	4,68	4,62	4,56	4,50	4,36	4,36
6	5,99	5,14	4,76	4,53	4,39	4,28	4,21	4,15	4,10	4,06	4,00	3,94	3,87	3,81	3,67	3,67
7	5,59	4,74	4,35	4,12	3,97	3,87	3,79	3,73	3,68	3,64	3,57	3,51	3,44	3,38	3,23	3,23
8	5,32	4,46	4,07	3,84	3,69	3,58	3,50	3,44	3,39	3,35	3,28	3,22	3,15	3,08	2,93	2,93
9	5,12	4,26	3,86	3,63	3,48	3,37	3,29	3,23	3,18	3,14	3,07	3,01	2,94	2,86	2,71	2,71
10	4,96	4,10	3,71	3,48	3,33	3,22	3,14	3,07	3,02	2,98	2,91	2,85	2,77	2,70	2,54	2,54
11	4,84	3,98	3,59	3,36	3,20	3,09	3,01	2,95	2,90	2,85	2,79	2,72	2,65	2,57	2,40	2,40
12	4,75	3,89	3,49	3,26	3,11	3,00	2,91	2,85	2,80	2,75	2,69	2,62	2,54	2,47	2,30	2,30
13	4,67	3,81	3,41	3,18	3,03	2,92	2,83	2,77	2,71	2,67	2,60	2,53	2,46	2,38	2,21	2,21
14	4,60	3,74	3,34	3,11	2,96	2,85	2,76	2,70	2,65	2,60	2,53	2,46	2,39	2,31	2,13	2,13
15	4,54	3,68	3,29	3,06	2,90	2,79	2,71	2,64	2,59	2,54	2,48	2,40	2,33	2,25	2,07	2,07

16	4,49	3,63	3,24	3,01	2,85	2,74	2,66	2,59	2,54	2,49	2,42	2,35	2,28	2,19	2,01
17	4,45	3,59	3,20	2,96	2,81	2,70	2,61	2,55	2,49	2,45	2,38	2,31	2,23	2,15	1,96
18	4,41	3,55	3,16	2,93	2,77	2,66	2,58	2,51	2,46	2,41	2,34	2,27	2,19	2,11	1,92
19	4,38	3,52	3,13	2,90	2,74	2,63	2,54	2,48	2,42	2,38	2,31	2,2	2,16	2,07	1,88
20	4,35	3,49	3,10	2,87	2,71	2,60	2,51	2,45	2,39	2,35	2,28	2,20	2,12	2,04	1,84
21	4,32	3,47	3,07	2,84	2,68	2,57	2,49	2,42	2,37	2,32	2,25	2,18	2,10	2,01	1,81
22	4,30	3,44	3,05	2,82	2,66	2,55	2,46	2,40	2,34	2,30	2,23	2,15	2,07	1,98	1,78
23	4,28	3,42	3,03	2,80	2,64	2,53	2,44	2,37	2,32	2,27	2,20	2,13	2,05	1,96	1,76
24	4,26	3,40	3,01	2,78	2,62	2,51	2,42	2,36	2,30	2,25	2,18	2,11	2,03	1,94	1,73
25	4,24	3,39	2,99	2,76	2,60	2,49	2,40	2,34	2,28	2,24	2,16	2,09	2,01	1,92	1,71
26	4,23	3,37	2,98	2,74	2,59	2,47	2,39	2,32	2,27	2,22	2,15	2,07	1,99	1,90	1,69
27	4,21	3,35	2,96	2,73	2,57	2,46	2,37	2,31	2,25	2,20	2,13	2,06	1,97	1,88	1,67
28	4,20	3,34	2,95	2,71	2,56	2,45	2,36	2,29	2,24	2,19	2,12	2,04	1,96	1,87	1,65
29	4,18	3,33	2,93	2,70	2,55	2,43	2,35	2,28	2,22	2,18	2,10	2,03	1,94	1,85	1,64
30	4,17	3,32	2,92	2,69	2,53	2,42	2,33	2,27	2,21	2,16	2,09	2,01	1,93	1,84	1,62
40	4,08	3,23	2,84	2,61	2,45	2,34	2,25	2,18	2,12	2,08	2,00	1,92	1,84	1,74	1,51
60	4,00	3,15	2,76	2,53	2,37	2,25	2,17	2,10	2,04	1,99	1,92	1,84	1,75	1,65	1,39
120	3,92	3,07	2,68	2,45	2,29	2,17	2,09	2,02	1,96	1,91	1,83	1,75	1,66	1,55	1,25
∞	3,84	3,00	2,60	2,37	2,21	2,10	2,01	1,94	1,88	1,83	1,75	1,67	1,57	1,46	1,00

Примечание. f_1 — число степеней свободы большей дисперсии; f_2 — число степеней свободы меньшей дисперсии.

Оглавление

Глава 1. ОСНОВНЫЕ ПОНЯТИЯ ТЕОРИИ ВЕРОЯТНОСТЕЙ И МАТЕМАТИЧЕСКОЙ СТАТИСТИКИ.....	3.
1.1. Случайная величина и ее закон распределения	3.
1.2. Простейшие приемы статистического описания	9.
1.3. Точечные и интервальные оценки	15
1.3.1. Выборочные характеристики	15
1.3.2. Доверительный интервал для математического ожидания	17
1.3.3. Доверительный интервал для среднеквадратического отклонения	19
1.4. Основные положения теории проверки статистических гипотез.....	22
1.4.1. Гипотеза в математической статистике.....	22
1.4.2. Гипотеза о равенстве двух дисперсий при неизвестных средних (критерий Фишера — Снедекора)	25
1.4.3. Гипотеза о равенстве средних двух нормальных распределений	27
1.4.4. Гипотеза о равенстве нескольких дисперсий при неизвестных средних (критерий Бартлетта)	30
1.4.5. Проверка гипотез о виде распределения	39
1.5. Корреляционно-регрессионный анализ.....	45
1.5.1. Понятие корреляционного анализа.....	45
1.5.2. Множественная корреляция	49.
1.5.3. Введение в регрессионный анализ.....	52
1.5.4. Нелинейная регрессия	56
1.5.5. Множественная регрессия	57.
Глава 2. МАТЕМАТИЧЕСКИЕ МЕТОДЫ В ТЕХНОЛОГИИ СТРОИТЕЛЬНЫХ МАТЕРИАЛОВ	59.
2.1. Цель планирования эксперимента.....	59.
2.2. Основные виды математических моделей	60
2.3. Метод наименьших квадратов.....	64.
2.3.1. Метод наименьших квадратов для моделей с одной переменной.....	64
2.3.2. Метод наименьших квадратов для многофакторных экспериментов.....	70
2.3.3. Статистический анализ уравнения регрессии.....	73
Глава 3. ЭКСПЕРИМЕНТАЛЬНЫЕ ПЛАНЫ ПЕРВОГО ПОРЯДКА	81
3.1. Этапы планирования эксперимента	81
3.2. Полный факторный план типа 2^k	82
3.3. Свойства полного факторного плана типа 2^k	85
3.4. Расчет коэффициентов регрессии.....	86.
3.5. Статистический анализ регрессионной модели, полученной по результатам ПФП 2^k	88
3.6. Применение ПФП 2^3	89.
3.7. Дробные факторные планы	96
Глава 4. ЭКСПЕРИМЕНТАЛЬНЫЕ ПЛАНЫ ВТОРОГО ПОРЯДКА	100
4.1. В-планы второго порядка	100
4.2. Применение В-плана второго порядка	105
4.2.1. Расчет уравнения регрессии.....	105
4.2.2. Анализ и интерпретация уравнения регрессии	113
4.3. Униформ-ротатабельный план	117
4.4. Исследование регрессионных моделей второго порядка для решения задач оптимизации	121
4.5. Композиционный план	124
4.5.1. Виды композиционных планов	124
4.5.2. Ортогональный композиционный план.....	125
4.6. Применение композиционного плана.....	132
Библиографический список	143
Приложение	144

Минимальные системные требования определяются соответствующими требованиями программ Adobe Reader версии не ниже 11-й либо Adobe Digital Editions версии не ниже 4.5 для платформ Windows, Mac OS, Android и iOS; экран 10"

Учебное электронное издание

**Александрова Ольга Владимировна,
Мацеевич Татьяна Анатольевна,
Кириянова Людмила Владимировна**

СТАТИСТИЧЕСКИЕ МЕТОДЫ РЕШЕНИЯ ТЕХНОЛОГИЧЕСКИХ ЗАДАЧ

Редактор *А.К. Смирнова*
Корректор *В.К. Чупрова*
Компьютерная правка *О.В. Суховой*
Верстка макета *О.Г. Горюновой*
Дизайн обложки *Д.Л. Разумного*

Подписано к использованию 25.08.2017
Формат 12,5×19,5 см
Гарнитура NewtonС

Федеральное государственное бюджетное образовательное
учреждение высшего образования
«Национальный исследовательский
Московский государственный строительный университет».
129337, Москва, Ярославское ш., 26.

Издательство МИСИ—МГСУ.
Тел.: (495) 287-49-14 (вн. 13-71), (499) 188-29-75, (499) 183-97-95
Сайт: <http://mgsu.ru>
Эл. почта: ric@mgsu.ru, rio@mgsu.ru

Электронная версия данной книги подготовлена
Агентством электронных изданий «Интермедиагор»
Сайт: <http://www.intermediator.ru>
Телефон: (495) 587-74-81
Эл. почта: info@intermediator.ru