

СПб ГУТ)))

Санкт-Петербургский государственный университет
телекоммуникаций им. проф. М.А. Бонч-Бруевича



Моделирование информационных систем

Санкт-Петербург

Моделирование систем

КУРС 4, СПЕЦИАЛЬНОСТЬ 230201.65

Лектор

Белов Михаил Петрович

**к.т.н., профессор кафедры Информационных
управляющих систем**

ОСНОВНЫЕ ПРИНЦИПЫ ПЛАНИРОВАНИЯ МАШИННОГО ЭКСПЕРИМЕНТА С МОДЕЛЬЮ

Для получения адекватной математической модели необходимо обеспечить выполнение определенных условий проведения эксперимента. Модель называют **адекватной**, если в оговоренной области варьирования факторов X полученные с помощью модели значения функций отклика Y отличаются от истинных не более чем на заданную величину.

Методы построения экспериментальных факторных моделей рассматриваются в **теории планирования эксперимента**.

Цель планирования эксперимента – получение максимума информации о свойствах исследуемого объекта при минимуме опытов. Такой подход обусловлен высокой стоимостью экспериментов, как физических, так и вычислительных, и вместе с тем необходимостью построения адекватной модели.

Планирование осуществляют как **активного**, так и **пассивного** эксперимента. Планируемый активный эксперимент при прочих равных условиях точнее и информативнее, а иногда и дешевле пассивного. Это следует учитывать при выборе вида эксперимента. В вычислительном эксперименте, в отличие от физического, нет никаких ограничений на выбор управляемых факторов и характер их изменения. Поэтому вычислительные эксперименты обычно всегда реализуются как активные. В дальнейшем будут рассматриваться в основном вопросы, связанные с планированием активных экспериментов.

При **планировании активных экспериментов** используются **следующие принципы**:

- отказ от полного перебора всех возможных состояний объекта;
- постепенное усложнение структуры математической модели;
- сопоставление результатов эксперимента с величиной случайных помех;
- рандомизация опытов;
- оптимальное планирование эксперимента.

Детальное представление о свойствах поверхности отклика может быть получено лишь при условии использования густой дискретной сетки значений факторов, покрывающей все факторное пространство. В узлах этой многомерной сетки находятся точки плана, в которых проводятся опыты. В этом случае в принципе можно получить факторную модель, которая будет практически почти полностью соответствовать исходной теоретической модели. Однако в большинстве случаев при решении практических задач, для которых используется факторная модель, такого детального описания не требуется. Выбор структуры факторной модели основан на постулировании определенной степени гладкости поверхности отклика. Поэтому с целью уменьшения количества опытов принимают небольшое число точек плана, для которых осуществляется реализация эксперимента.

В отсутствие априорной информации о свойствах функции отклика нет смысла сразу строить сложную математическую модель объекта. Если проверка этой модели на адекватность не дает удовлетворительного результата, ее постепенно усложняют путем изменения структуры (например, повышая степень полинома, принятого в качестве факторной модели, или вводя в модель до-полнительные факторы и т.п.).

При этом используются результаты опытов, выполненных при построении простой модели, и проводится некоторое количество дополнительных опытов.

При большом уровне случайной помехи получается большой разброс значений функции отклика Y в опытах, проведенных в одной и той же точке плана. В этом случае оказывается, что чем выше уровень помехи, тем с большей вероятностью простая модель окажется работоспособной. Чем меньше уровень помехи, тем точнее должна быть факторная модель.

Кроме случайной помехи при проведении эксперимента может иметь место **систематическая помеха**. Наличие этой помехи практически никак не обнаруживается и результат ее воздействия на функцию не поддается контролю. Однако если путем соответствующей организации проведения опытов искусственно создать случайную ситуацию, то систематическую помеху можно перевести в разряд случайных. Такой принцип организации эксперимента называют **рандомизацией** систематически действующих помех.

Наличие помех приводит к ошибкам эксперимента. **Ошибки**

подразделяют на **систематические** и **случайные**, соответственно наименованиям вызывающих их факторов – помех.

В вычислительных активных экспериментах ошибки характерны только для определяемых значений функций отклика. Если исходить из целей построения факторных моделей на основе теоретических моделей, полагая, что теоретические модели дают точное описание физических свойств технического объекта, а регрессионная модель является ее аппроксимацией, то значения функций отклика будут содержать только случайную ошибку, а этом случае необходимости в рандомизации опытов не возникает.

Рандомизацию опытов осуществляют только в физических экспериментах. Следует отметить, что в этих экспериментах систематическую ошибку может породить наряду с отмеченными в предыдущем параграфе факторами также неточное задание значений управляемых факторов, обусловленное некачественной калибровкой приборов для их измерения (инструментальная ошибка), конструктивными или технологическими факторами.

К факторам в активном эксперименте предъявляются **определенные требования**.

Они должны быть:

1) управляемыми (установка заданных значений и поддержание постоянными в процессе опыта);

2) совместными (их взаимное влияние не должно нарушать процесс функционирования объекта);

3) независимыми (уровень любого фактора должен устанавливаться независимо от уровней остальных);

4) однозначными (одни факторы не должны быть функцией других);

5) непосредственно влияющими на выходные параметры.

В вычислительном эксперименте реализация трех первых требований не создает никаких затруднений, а в физическом эксперименте могут возникнуть сложности и даже невозможность их осуществления, что приведет к необходимости замены активного эксперимента пассивным.

Функции отклика должны быть:

- **численно измеряемыми**;

- **иметь четкий физический смысл**;

- **однозначными** (характеризовать только одно свойство объекта);

- **информативными** (полностью характеризовать определенное свойство объекта);

- **статистически эффективными** (измеряться с достаточной точностью с целью сокращения дублирования опытов).

ПЛАН ЭКСПЕРИМЕНТА

При проведении активного эксперимента задается определенный план варьирования факторов, т.е. эксперимент заранее планируется.

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

Планирование эксперимента – выбор плана эксперимента, удовлетворяющего заданным требованиям.

Точка плана – упорядоченная совокупность численных значений факторов, соответствующая условиям проведения опыта, т.е. точка факторного пространства, в которой проводится эксперимент.

$$\bar{X}_i = (X_{i1}, X_{i2}, \dots, X_{in}).$$

Общая совокупность таких векторов $\bar{X}_i, i = \overline{1, L}$, образует план эксперимента, а совокупность различающихся векторов, число которых обозначим N , — *спектр плана*.

В активном эксперименте факторы могут принимать только фиксированные значения. Фиксированное значение фактора называют *уровнем фактора*. Количество принимаемых уровней факторов зависит от выбранной структуры факторной модели и принятого плана эксперимента. Минимальный $X_{j \min}$ и максимальный $X_{j \max}, j = \overline{1, n}$ (n — число факторов), уровни всех факторов выделяют в факторном пространстве некоторый гиперпараллелепипед, представляющий собой *область планирования*. В области планирования находятся все возможные значения факторов, используемые в эксперименте.

Вектор $X^0 = (X_1^0, X_2^0, \dots, X_n^0)$ задает точку центра области планирования. Координаты этой точки X_j^0 обычно выбирают из соотношения

$$X_j^0 = (X_{j \max} + X_{j \min}) / 2. \quad (11.1)$$

Точку \bar{X}^0 называют *центром эксперимента*. Она определяет основной уровень факторов $X_j^0, j = \overline{1, n}$. Центр эксперимента стремятся выбрать как можно ближе к точке, которая соответствует искомым оптимальным значениям факторов. Для этого используется априорная информация об объекте.

Интервалом (или шагом) варьирования фактора X_j называют величину, вычисляемую по формуле

$$\Delta X_j = (X_{j\max} - X_{j\min}) / 2, j = \overline{1, n}. \quad (11.2)$$

Факторы нормируют, а их уровни кодируют. В кодированном виде верхний уровень обозначают +1, нижний -1, а основной 0. Нормирование факторов осуществляют на основе соотношения

$$x_j = (X_j - X_j^0) / \Delta X_j, j = \overline{1, n}. \quad (11.3)$$

Для переменных x_j начало координат совмещено с центром эксперимента, а в качестве единиц измерения используются интервалы варьирования факторов. Геометрическое представление области планирования при двух факторах показано на рис. 11.2. Центр эксперимента находится в точке 0 с координатами X_1^0 , X_2^0 . Точки 1, 2, 3, 4 являются точками плана эксперимента. Например, значения факторов X_1 и X_2 в точке 1 равны соответственно $X_{1\min}$ и $X_{2\min}$, а нормированные их значения $x_{1\min} = -1$, $x_{2\min} = -1$.

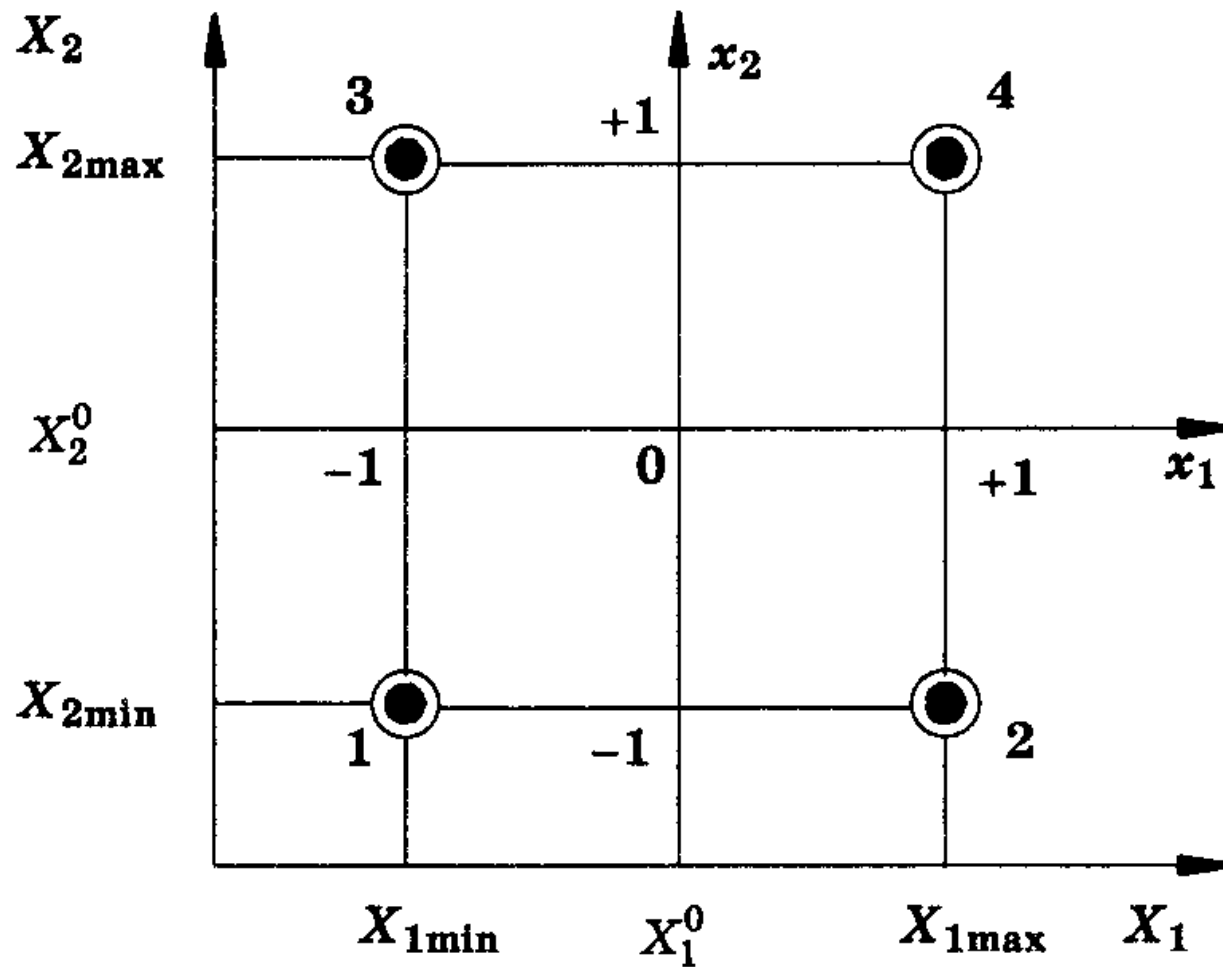


Рис. 2. Геометрическое представление области планирования при двух факторах: X_1 и X_2

В дальнейшем будем предполагать, что в планах активных экспериментов факторы нормированы.

План эксперимента удобно представлять в матричной форме. План эксперимента задается либо матрицей плана, либо матрицей спектра плана в совокупности с матрицей дублирования.

Матрица плана представляет собой прямоугольную таблицу, содержащую информацию о количестве и условиях проведения опытов. Строки матрицы плана соответствуют опытам, а столбцы – факторам. Размерность матрицы плана $L \times p$, где L – число опытов, p – число факторов. При проведении повторных (дублирующих) опытов в одних и тех же точках плана матрица плана содержит ряд совпадающих строк.

Матрица спектра плана — это матрица, в которую входят только различающиеся между собой строки матрицы плана. Размерность матрицы спектра плана $N \times n$, где N — число точек плана, различающихся между собой хотя бы одной координатой X_{ij} , $i = \overline{1, N}$, $j = \overline{1, n}$.

Матрица спектра плана имеет вид

$$X = \begin{bmatrix} \vec{X}_1 \\ \vec{X}_2 \\ \dots \\ \vec{X}_i \\ \dots \\ \vec{X}_N \end{bmatrix} = \begin{bmatrix} X_{11} & X_{12} & \dots & X_{1j} & \dots & X_{1n} \\ X_{21} & X_{22} & \dots & X_{2j} & \dots & X_{2n} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ X_{i1} & X_{i2} & \dots & X_{ij} & \dots & X_{in} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ X_{N1} & X_{N2} & \dots & X_{Nj} & \dots & X_{Nn} \end{bmatrix}, \quad (11.4)$$

где \bar{X}_i — вектор, определяющий нормированные значения координат точки плана в i -м опыте; X_{ij} — нормированное значение j -го фактора в i -м опыте.

Матрица дублирования — квадратная диагональная матрица m , диагональные элементы которой равны числам параллельных опытов в соответствующих точках спектра плана:

$$m = \begin{bmatrix} m_1 & 0 & \dots & 0 \\ 0 & m_2 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & m_N \end{bmatrix}. \quad (11.5)$$

Опыты при выполнении эксперимента проводятся в последовательности, предусмотренной матрицей плана. Эта матрица составляется лишь при необходимости рандомизации опытов, когда в результатах эксперимента можно ожидать наличие систематических ошибок. Для выбора случайной последовательности опытов используется таблица равномерно распределенных случайных чисел. Первое число таблицы выбирают произвольно, желательно случайным образом, а затем, начиная с этого числа, выписывают L чисел таблицы, где L — число опытов (с учетом их дублирования). При этом числа, большие L , а также уже выписанные, отбрасываются.

В вычислительных экспериментах опыты проводят в соответствии с матрицей спектра плана, так как предполагается отсутствие систематических ошибок и поэтому нет необходимости в рандомизации опытов.

РЕГРЕССИОННЫЙ АНАЛИЗ

Регрессионный анализ проводится с целью получения по экспериментальным данным регрессионных моделей, представляющих собой экспериментальные факторные модели. Задачей регрессионного анализа является определение параметров экспериментальных факторных моделей объектов проектирования или исследования, т.е. определение коэффициентов уравнений моделей при выбранной их структуре.

Регрессионный анализ включает **три основных этапа**:

- 1) статистический анализ результатов эксперимента;
- 2) получение коэффициентов регрессионной модели;
- 3) оценку адекватности и работоспособности полученной экспериментальной факторной модели технической системы.

Поскольку параметры фактических моделей b определяют по результатам ограниченного количества опытов, то получаемые их значения являются оценками истинных коэффициентов регрессии β .

Под структурой экспериментальной факторной математической модели понимается вид математических соотношений между факторами X, Z и откликом Y . В качестве факторов принимают внутренние и внешние параметры технической системы, подлежащие оптимизации в процессе ее проектирования. Внутренние параметры системы – это параметры ее элементов, внешние – это параметры внешней среды, в условиях воздействий которой осуществляется функционирование системы. Функциями отклика Y являются выходные параметры технической системы, характеризующие ее эффективность и качество процессов функционирования. Выходные параметры системы принимаются в качестве критериев оптимальности.

Как уже отмечалось, структура факторной модели выбирается на основе априорной информации, используя принцип постепенного ее усложнения. Параметры факторной математической модели определяются методами регрессионного анализа. При определении параметров этими методами нет необходимости различать виды факторов, т.е. подразделять факторы на управляемые \vec{X} и неуправляемые \vec{Z} . Поэтому в дальнейшем все они будут обозначаться буквой \vec{X} . Тогда факторную модель можно представить векторным уравнением регрессии вида

$$\vec{Y} = \vec{\varphi}(\vec{X}, \vec{b}). \quad (11.6)$$

Определение параметров \vec{b} этой модели будем рассматривать на примере одного уравнения $Y = \varphi(\vec{X}, \vec{b})$. Для определения параметров используются результаты эксперимента. Результаты эксперимента можно представить функцией вида

$$Y = \varphi(\vec{X}) + \varepsilon, \quad (11.7)$$

где ε — аддитивная помеха случайного характера с нормальным законом распределения.

Так как каждый опыт проводится при определенном сочетании уровней факторов \vec{X} , то функцию $\varphi(\vec{X})$ представим выражением:

$$\varphi(\vec{X}) = \sum_{j=0}^d \beta_j f_j(\vec{X}), \quad (11.8)$$

где β_j — j -й элемент вектора искомых коэффициентов уравнения регрессии: $\vec{\beta} = (\beta_0, \beta_1, \dots, \beta_d)^T$; $f_j(\vec{X})$ — j -я базисная функция — элемент вектора базисных функций $\vec{f}(\vec{X}) = [f_0(\vec{X}), f_1(\vec{X}), \dots, f_d(\vec{X})]^T$.

В качестве базисных функций используют переменные простейших полиномов, системы ортогональных полиномов (Эрмита, Лежандра, Лаггера и др.), тригонометрические функции. Наиболее часто пользуются простейшими полиномами первой и второй степеней. Например, полином первой степени, описывающий функцию отклика y при двух факторах x_1 и x_2 , может иметь вид

$$y = b_0 + b_1x_1 + b_2x_2, \quad (11.9)$$

или

$$y = b_0 + b_1x_1 + b_2x_2 + b_3x_1x_2, \quad (11.10)$$

а полином второй степени

$$y = b_0 + b_1x_1 + b_2x_2 + b_3x_1x_2 + b_4x_1^2 + b_5x_2^2. \quad (11.11)$$

Базисные функции в случае использования последнего выражения имеют вид: $f_0(\vec{X}) = 1$; $f_1(\vec{X}) = x_1$; $f_2(\vec{X}) = x_2$; $f_3(\vec{X}) = x_1x_2$; $f_4(\vec{X}) = x_1^2$; $f_5(\vec{X}) = x_2^2$.

Если уравнение регрессии имеет вид выражений (11.9), (11.10), его называют уравнением **линейной регрессии** (линейной регрессией или регрессией первого порядка), а если содержит факторы во второй и более высокой степени – **нелинейной регрессией** (регрессией соответствующего порядка).

Линейная регрессия может представлять как линейную математическую модель, так и нелинейную, в зависимости от того, содержит ли она **линейные эффекты** (как в выражении (11.9)), или наряду с ними также **эффекты взаимодействия** (как в выражении (11.10)). Линейным называют эффект, характеризующий линейную зависимость выходного параметра y от соответствующего фактора x_i . Эффектом взаимодействия называют эффект, характеризующий совместное влияние нескольких факторов на y (например, в выражении (11.10) x_1x_2). Эффекты взаимодействия двух факторов называют парным взаимодействием, трех факторов – тройным взаимодействием и т.д.

Как всякий статистический метод, регрессионный анализ применим при определенных предпосылках (постулатах).

1. Аддитивная помеха ε – случайная нормально распределенная величина с параметрами $m_\varepsilon = 0$ и $\sigma^2_\varepsilon = \text{const}$. В этом случае функция отклика Y также случайная величина с нормальным законом распределения. Гипотезу о нормальном распределении Y можно проверить по критерию Пирсона.

2. Постоянство дисперсии помехи означает, что интенсивность ошибки определения Y не меняется при изменении уровня факторов в процессе эксперимента. Выполнение этого постулата проверяется по критерию однородности дисперсии в разных точках спектра плана.

3. Значения факторов в активном эксперименте – неслучайные величины. Это означает, что установление каждого фактора на заданном уровне и удерживание его на этом уровне во время опыта точнее, чем ошибка воспроизводимости. В вычислительном эксперименте это выполняется однозначно, а в физическом вклад, вносимый ошибками измерения факторов X , должен быть пренебрежимо малым в сравнении с действием других неконтролируемых факторов, образующих ошибку ε определения функции Y .

Значения помехи ε в различных точках опыта некоррелированы. Для обеспечения этого требования используется рандомизация опытов.

В пассивном эксперименте условие некоррелированности помехи обеспечиваются путем соответствующего выбора временного интервала съема информации об условиях и результатах опытов.

Векторы-столбцы базисных функций должны быть линейно независимыми. Выполнение этого требования необходимо для получения отдельных оценок b всех коэффициентов регрессии β . В активном эксперименте оно обеспечивается соответствующим выбором спектра плана эксперимента. При этом число опытов N (без учета дублирования) должно быть не меньше, чем число оцениваемых коэффициентов N_B , т.е. $N \geq N_B$.

В пассивном эксперименте линейная зависимость между столбцами практически исключается, так как факторы неуправляемы и принимают случайные значения в разных опытах, но может наблюдаться сильная коррелированность столбцов, что повлечет за собой большие ошибки вычисления коэффициентов регрессии. Для выявления коррелированности столбцов проводят корреляционный анализ результатов пассивного эксперимента.

ОЦЕНКА ПАРАМЕТРОВ РЕГРЕССИОННОЙ МОДЕЛИ

Исходными данными для получения оценок параметров регрессионной модели технической системы (т.е. оценок \vec{b} искоемых коэффициентов регрессии $\vec{\beta}$) является информация о значениях управляемых факторов \vec{X} (или неуправляемых — при проведении пассивного эксперимента) и функции отклика Y . Эту информацию можно представить в виде матрицы X значений факторов во всех N опытах, предусмотренных спектром плана эксперимента, и вектора-столбца \vec{Y} полученных в этих опытах значений функции отклика Y :

$$X = \begin{bmatrix} \vec{X}_1 \\ \vec{X}_2 \\ \dots \\ \vec{X}_i \\ \dots \\ \vec{X}_N \end{bmatrix} = \begin{bmatrix} X_{11} & X_{12} & \dots & X_{1j} & \dots & X_{1n} \\ X_{21} & X_{22} & \dots & X_{2j} & \dots & X_{2n} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ X_{i1} & X_{i2} & \dots & X_{ij} & \dots & X_{in} \\ \dots & \dots & \dots & \dots & \dots & \dots \\ X_{N1} & X_{N2} & \dots & X_{Nj} & \dots & X_{Nn} \end{bmatrix}; \quad (11.12)$$

$$\vec{Y} = (y_1, y_2, \dots, y_i, \dots, y_N)^T, \quad (11.13)$$

где $X_i = (X_{i1}, X_{i2}, \dots, X_{in})$ — вектор-строка значений факторов в i -м опыте; X_{ij} — значение j -го фактора в i -м опыте; n — количество факторов; N — количество опытов; y_i — значение функции отклика Y в i -м опыте (если проводились параллельные опыты, т.е. осуществлялось дублирование опытов, то вместо y_i используются оценки их математических ожиданий, т.е. выборочные средние \bar{y}_i).

Значения базисных функций во всех опытах представляют собой матрицу F , называемую *матрицей базисных функций*

$$F = \begin{bmatrix} \bar{f}_1(\bar{X}_1) \\ \bar{f}_2(\bar{X}_2) \\ \dots \\ \bar{f}_i(\bar{X}_i) \\ \dots \\ \bar{f}_N(\bar{X}_N) \end{bmatrix} = \begin{bmatrix} f_{10} & f_{11} & f_{12} & \dots & f_{1k} & \dots & f_{1d} \\ f_{20} & f_{21} & f_{22} & \dots & f_{2k} & \dots & f_{2d} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ f_{i0} & f_{i1} & f_{i2} & \dots & f_{ik} & \dots & f_{id} \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ f_{N0} & f_{N1} & f_{N2} & \dots & f_{Nk} & \dots & f_{Nd} \end{bmatrix}, \quad (11.14)$$

где f_{ik} — значение k -й базисной функции в i -м опыте; $\bar{f}_i(\bar{X}_i) =$

$= (f_{i0}, f_{i1}, f_{i2}, \dots, f_{ik}, \dots, f_{id})$ — вектор-строка значений базисных функций в i -м опыте.

Используя информацию об X , \bar{Y} и F , необходимо найти оценки коэффициентов регрессии, представляемые вектором-столбцом

$$\bar{b}^T = (b_0, b_1, b_2, \dots, b_k, \dots, b_d), \quad (11.15)$$

где b_k — значение оценки коэффициента регрессии при базисной функции $f_k(\bar{X})$.

Так как функция отклика Y — случайная величина, поскольку на ее значения в различных опытах оказывает влияние случайная помеха ε , то оценки коэффициентов регрессии будут случайными величинами.

Уравнение регрессии устанавливает зависимость между оценкой математического ожидания функции отклика \bar{y} и факторами $\bar{X} = (x_1, x_2, \dots, x_n)$. Общий вид этой зависимости

$$\bar{y} = \sum_{k=0}^d b_k f_k(\bar{X}). \quad (11.16)$$

В связи с наличием помехи значение функции отклика в i -м опыте y_i будет отличаться от \bar{y}_i . Для определения y_i можно составить выражение:

$$y_i = b_0 f_{i0} + b_1 f_{i1} + \dots + b_k f_{ik} + \dots + b_d f_{id} + \varepsilon_i, i = \overline{1, N}, \quad (11.17)$$

где ε_i — невязка уравнения регрессии в i -м опыте.

Невязка характеризует отклонение значений функции отклика в опытах от получаемых с помощью регрессионной модели (11.16). Она возникает по двум причинам: из-за ошибки эксперимента и из-за непригодности (приближенности) выбранной структуры факторной математической модели. Причем эти причины смешаны и нельзя сказать, какая из них преобладает.

Если постулировать, что модель пригодна, то невязка будет порождаться только ошибкой опыта. Тогда для определения коэффициентов уравнения (11.16) невязку надо минимизировать. Для этого в регрессионном анализе используется *метод наименьших квадратов* (МНК). Составляется функция, представляющая собой сумму квадратов невязок, и осуществляется ее минимизация:

$$E = \sum_{i=1}^N \varepsilon_i^2 \rightarrow \min. \quad (11.18)$$

Подставим значение ε_i из выражения (11.17):

$$E = \sum_{i=1}^N [y_i - (b_0 f_{i0} + b_1 f_{i1} + \dots + b_k f_{ik} + \dots + b_d f_{id})]^2 \rightarrow \min. \quad (11.19)$$

В выражении (11.19) коэффициенты b_k рассматриваются как неизвестные переменные, которые наилучшим образом соответствуют полученным результатам эксперимента. Значения этих коэффициентов, при которых достигается минимум функции E , принимаются в качестве оценок коэффициентов регрессии. Минимум функции E имеет место при равенстве нулю частных производных этой функции по переменным b_0, b_1, \dots, b_d :

$$\frac{\partial E}{\partial b_0} = -2 \sum_{i=1}^N [y_i - (b_0 f_{i0} + b_1 f_{i1} + \dots + b_d f_{id})] f_{i0} = 0;$$

$$\frac{\partial E}{\partial b_1} = -2 \sum_{i=1}^N [y_i - (b_0 f_{i0} + b_1 f_{i1} + \dots + b_d f_{id})] f_{i1} = 0;$$

.....

$$\frac{\partial E}{\partial b_d} = -2 \sum_{i=1}^N [y_i - (b_0 f_{i0} + b_1 f_{i1} + \dots + b_d f_{id})] f_{id} = 0.$$

$$\Phi = \begin{bmatrix} \sum_{i=1}^N f_{i0}^2 & \sum_{i=1}^N f_{i1}f_{i0} & \dots & \sum_{i=1}^N f_{id}f_{i0} \\ \sum_{i=1}^N f_{i0}f_{i1} & \sum_{i=1}^N f_{i1}^2 & \dots & \sum_{i=1}^N f_{id}f_{i1} \\ \dots & \dots & \dots & \dots \\ \sum_{i=1}^N f_{i0}f_{id} & \sum_{i=1}^N f_{i1}f_{id} & \dots & \sum_{i=1}^N f_{id}^2 \end{bmatrix}. \quad (11.22)$$

Матрицу Φ называют *информационной матрицей Фишера*. Она содержит $(d + 1)$ строк и $(d + 1)$ столбцов, причем элемент j -й строки k -го столбца представляет собой сумму $\sum_{i=1}^N f_{ij}f_{ik}$. Матрица

Φ симметрична относительно главной диагонали, что упрощает составление системы алгебраических уравнений (11.20) для регрессионной модели.

Систему уравнений (11.20) можно также записать в матричной форме

$$\Phi \vec{b} = F^T \vec{Y}. \quad (11.23)$$

Система уравнений (11.20) имеет единственное решение, если определитель матрицы Φ не равен нулю. В этом случае матрица Φ будет не вырожденной. Выполнение пятой предпосылки регрессионного анализа, изложенной в предыдущем параграфе, исключает возникновение вырожденности.

методом Гаусса. При небольшом числе определяемых коэффициентов b_k можно использовать правило Крамера.

Полученные методом наименьших квадратов оценки b_0, b_1, \dots, b_d действительных значений коэффициентов регрессии $\beta_0, \beta_1, \dots, \beta_d$ обладают следующими свойствами:

1) математические ожидания оценок $M[b_j] = \beta_j, j = \overline{0, d}$, т.е.

оценки b_j несмещенные;

2) дисперсии оценок коэффициентов регрессии минимальны и равны

$$\sigma_{b_j}^2 = M\left\{\left(b_j - M[b_j]\right)^2\right\} = M\left\{\left(b_j - \beta_j\right)^2\right\} = \sigma_{\varepsilon}^2 C_{jj}, \quad (11.24)$$

а корреляционный момент

$$\begin{aligned}\mu_{11}(b_j, b_k) &= M\left\{\left(b_j - M[b_j]\right)\left(b_k - M[b_k]\right)\right\} = \\ &= M\left\{\left(b_j - \beta_j\right)\left(b_k - \beta_k\right)\right\} = \sigma_\varepsilon^2 C_{jk},\end{aligned}\tag{11.25}$$

где C_{jj} , C_{jk} — элементы матрицы Φ^{-1} , обратной к информационной; σ_ε^2 — дисперсия случайной помехи;

3) оценки b_0, b_1, \dots, b_d подчиняются совместному $(d + 1)$ -мерному нормальному распределению.

ПЛАНЫ ЭКСПЕРИМЕНТОВ И ИХ СВОЙСТВА

Для проведения активных экспериментов разработано множество различных планов. Планы учитывают как особенности структуры регрессионных моделей, так и требования их эффективности с позиций повышения точности получаемых моделей и снижения затрат на проведение эксперимента.

При построении линейных моделей или нелинейных, содержащих только взаимодействия факторов, но без квадратов этих факторов (**регрессий первого порядка**), каждый фактор можно варьировать только на двух уровнях. Для получения таких моделей используют **планы первого порядка**.

Известно несколько разновидностей планов первого порядка. Эти планы различаются в зависимости от структуры регрессионной модели.

Они предназначены для планирования следующих видов экспериментов:

- **однофакторного (классического) эксперимента;**
- **полного факторного эксперимента;**
- **дробного факторного эксперимента.**

Если в регрессионную модель входят факторы в квадрате или с более высокими степенями, то необходимо не менее трех уровней варьирования факторов. При построении квадратичных моделей применяют **планы второго порядка**. Эти планы часто используют в качестве своего ядра какой-либо план первого порядка, который дополняется так называемыми **звездными точками**.

Планы различают по степени насыщенности и композиционности. **План называют насыщенным**, если общее число точек плана равно числу неизвестных параметров регрессионной модели. Такой план позволяет получить экспериментальную факторную модель при минимальных затратах, так как обеспечивает минимум числа опытов.

План называется композиционным, если в его спектр в качестве составной части входят точки спектра плана, который был реализован при построении более простой модели. Композиционность плана позволяет реализовать принцип постепенного усложнения модели при минимальных затратах, так как при этом используются результаты опытов, выполненных для получения простой модели. Многие планы второго порядка являются композиционными.

Важным свойством плана является его *ортогональность*. У ортогональных планов информационная матрица Фишера Φ диагональная, а столбцы матрицы базисных функций F попарно ортогональны. Для ортогонального плана при заданных значениях диагональных элементов матрицы Φ дисперсии $\sigma_{b_k}^2$ оценок коэффициентов регрессии b_k минимальны. Причем, эти оценки получаются независимыми, что существенно облегчает их вычисление и анализ.

При изменении вида плана изменяется матрица Φ , что влияет на дисперсии оценок коэффициентов регрессии. Различают *D*-, *A*- и *E*-оптимальные планы. Они обеспечивают различные формы эллипсоидов рассеивания оценок. *D*-оптимальный план минимизирует обобщенную дисперсию оценок коэффициентов регрессии и обеспечивает минимальный объем эллипсоида их рассеивания. *A*-оптимальный план минимизирует среднюю дисперсию всех оценок, а эллипсоид имеет наименьшую сумму квадратов длин осей. Эллипсоид рассеивания у *E*-оптимального плана имеет минимальную длину своей наибольшей оси.

В зависимости от возможностей предсказания отклика по уравнению регрессии различают планы **ротатабельные** и **униформные**. План называется **ротатабельным**, если дисперсия предсказания отклика постоянна на фиксированном расстоянии от центра эксперимента. **Униформный план** обеспечивает практически постоянное ее значение в некоторой области факторного пространства. Свойства **ротатабельности** или **униформности** обеспечиваются соответствующим выбором точек матрицы спектра плана. Задача выбора оптимального плана довольно сложная и в большинстве случаев не имеет аналитического решения. Поэтому поиск оптимальных планов обычно осуществляется численными методами на ЭВМ.

Рассмотрим основы построения и основные свойства планов первого порядка.

ПЛАН ОДНОФАКТОРНОГО ЭКСПЕРИМЕНТА

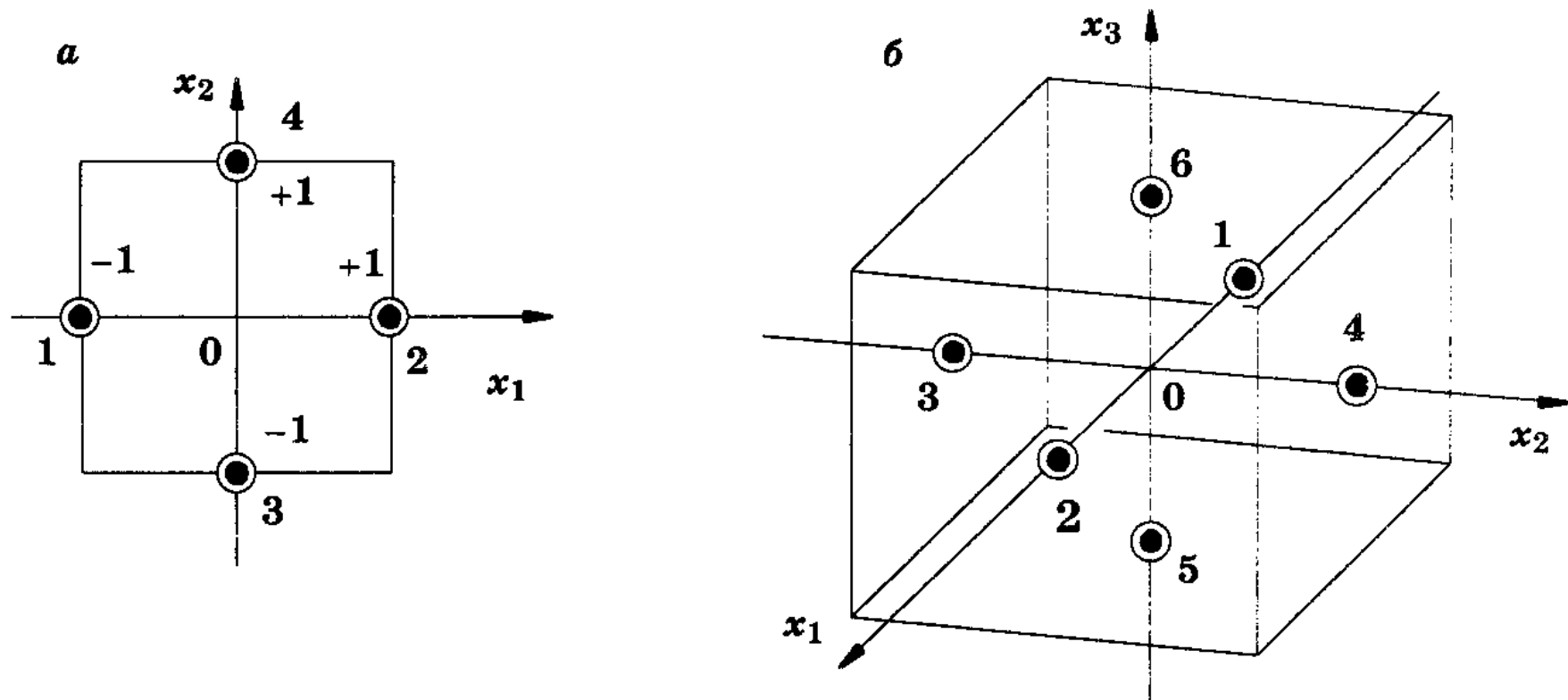
Однофакторный (классический) эксперимент предназначен для получения линейной экспериментальной факторной модели вида

$$y = b_0 + b_1x_1 + b_2x_2 + \dots + b_nx_n. \quad (11.26)$$

Однофакторный эксперимент предусматривает поочередное варьирование каждого из факторов при фиксированных на некотором уровне значениях остальных факторов. Фактор X_i варьируют на двух уровнях $X_{iВ}$ и $X_{iН}$, а все остальные при этом должны находиться в точке центра эксперимента $X_j^0, j \neq i$. Для нормированных факторов $x_{iВ} = +1, x_{iН} = -1, x_j = 0$. С учетом этого составим матрицу спектра плана однофакторного эксперимента

$$X = \begin{bmatrix} -1 & 0 & \dots & 0 \\ +1 & 0 & \dots & 0 \\ 0 & -1 & \dots & 0 \\ 0 & +1 & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & -1 \\ 0 & 0 & \dots & +1 \end{bmatrix}. \quad (11.27)$$

Число точек плана в этом случае $N = 2n$, где n — количество факторов. Точки спектра плана располагаются в центрах граней гиперкуба. На рис. 11.3, *a* показано расположение точек для двумерного случая, а на рис. 11.3, *б* — для трехмерного.



**Рис. 11.3. Расположение точек спектра
плана однофакторного эксперимента:
a — при двух факторах; *б* — при трех факторах**

Вектор базисных функций имеет вид

$$\vec{f}(\vec{X}) = (1, x_1, x_2, \dots, x_n), \quad (11.28)$$

а матрица F численных значений базисных функций отличается от матрицы спектра плана X только одним дополнительным столбцом, соответствующим базисной функции $f_0 = (\vec{X}) = 1$,

$$F = \begin{bmatrix} +1 & -1 & 0 & \dots & 0 \\ +1 & +1 & 0 & \dots & 0 \\ +1 & 0 & -1 & \dots & 0 \\ +1 & 0 & +1 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ +1 & 0 & 0 & \dots & -1 \\ +1 & 0 & 0 & \dots & +1 \end{bmatrix}. \quad (11.29)$$

Матрица базисных функций F обладает очевидными свойствами:

$$1) \quad \sum_{i=1}^N f_k(\vec{X}_i) = 0, \quad k = \overline{1, n}; \quad (11.30)$$

$$\sum_{i=1}^N f_0(\vec{X}_i) = N = 2n; \quad (11.31)$$

$$2) \quad \sum_{i=1}^N [f_k(\vec{X}_i)]^2 = 2, \quad k = \overline{1, n}; \quad (11.32)$$

$$\sum_{i=1}^N [f_0(\vec{X}_i)]^2 = N = 2n; \quad (11.33)$$

$$3) \quad \sum_{i=1}^N f_j(\vec{X}_i) f_k(\vec{X}_i) = 0, \quad j \neq k; \quad j, k = \overline{0, n}, \quad (11.34)$$

где N — число точек спектра плана; $f_k(\vec{X}_i)$ — значение k -й базисной функции в i -м опыте.

Согласно выражению (11.34) векторы-столбцы всех базисных функций попарно ортогональны.

Используя свойства (11.32) — (11.34) и выражение (11.22), легко составить информационную матрицу Фишера $\Phi = F^T F$:

$$\Phi = \begin{bmatrix} 2n & 0 & 0 & \dots & 0 \\ 0 & 2 & 0 & \dots & 0 \\ 0 & 0 & 2 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & 2 \end{bmatrix}. \quad (11.35)$$

Так как матрица Φ диагональная, то план однофакторного эксперимента ортогональный и коэффициенты регрессии некоррелированы друг с другом. Для определения дисперсии оценок коэффициентов регрессии (11.26) вычислим обращенную матрицу Фишера

$$\Phi^{-1} = \begin{bmatrix} 1/2n & 0 & 0 & \dots & 0 \\ 0 & 1/2 & 0 & \dots & 0 \\ 0 & 0 & 1/2 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & 1/2 \end{bmatrix}. \quad (11.36)$$

Искомые дисперсии оценок коэффициентов регрессии определяются произведениями дисперсии помехи σ_{ε}^2 на соответствующие диагональные элементы матрицы Φ^{-1} :

$$\sigma_{b_0}^2 = \sigma_{\varepsilon}^2 / (2n); \sigma_{b_k}^2 = \sigma_{\varepsilon}^2 / 2, k = \overline{1, n}. \quad (11.37)$$

Очевидно, что точность получаемой модели в этом случае невысокая, так как коэффициенты регрессии $b_k, k = \overline{1, n}$ (кроме коэффициента b_0), имеют высокое значение дисперсии. Поэтому однофакторный эксперимент следует признать явно неудовлетворительным для построения модели технической системы. В связи с этим в настоящее время он практически не применяется. Следует отметить, что рассмотренный план обладает свойством ротативности.

ПЛАН ПОЛНОГО ФАКТОРНОГО ЭКСПЕРИМЕНТА

Спектр плана полного факторного эксперимента (ПФЭ) содержит все возможные комбинации значений факторов на всех уровнях их изменения. Число точек N спектра плана определяется по формуле

$$N = U^n, \quad (11.38)$$

где U — число уровней варьирования факторов; n — количество факторов.

Рассмотрим особенности и свойства ПФЭ, применяемых при построении линейных регрессий вида

$$\begin{aligned}
 y = & b_0 + \sum_{j=1}^n b_j x_j + \sum_{j=1}^n \sum_{k=j+1}^n b_{j,k} x_j x_k + \\
 & + \sum_{j=1}^n \sum_{k=j+1}^n \sum_{l=k+1}^n b_{j,k,l} x_j x_k x_l + \dots + b_{1,2,\dots,n} x_1 x_2 \dots x_n.
 \end{aligned}
 \tag{11.39}$$

Для получения линейной регрессии достаточно варьировать факторы на двух уровнях, т.е. $U = 2$. Тогда число точек спектра плана

$$N = 2^n. \tag{11.40}$$

Такой план принято обозначать ПФЭ 2^n .

Рассмотрим порядок составления матрицы спектра плана, полагая, что факторы нормированы и, следовательно, могут принимать значения только либо $+1$, либо -1 . Напомним, что столбцы матрицы X соответствуют значениям факторов x_1, x_2, \dots, x_n .

Для составления матрицы спектра плана используется следующее простое правило: в первой строке матрицы все факторы равны -1 , в первом столбце знаки единиц меняются поочередно; во втором столбце они чередуются через два; в третьем — через 4; в четвертом — через 8 и т.д. по степеням двойки. Следовательно, для каждого последующего столбца частота изменения знака в 2 раза меньше, чем для предыдущего.

Используя изложенное правило чередования знаков, составим матрицы спектров планов для случаев $n = 2$ и $n = 3$, т.е. для двух и трех факторов.

При $n = 2$ число точек плана $N = 2^2 = 4$, а матрица спектра плана имеет вид

$$X = \begin{bmatrix} -1 & -1 \\ +1 & -1 \\ -1 & +1 \\ +1 & +1 \end{bmatrix}, \quad (11.41)$$

при $n = 3$ $N = 2^3 = 8$, а матрица X

$$X = \begin{bmatrix} -1 & -1 & -1 \\ +1 & -1 & -1 \\ -1 & +1 & -1 \\ +1 & +1 & -1 \\ -1 & -1 & +1 \\ +1 & -1 & +1 \\ -1 & +1 & +1 \\ +1 & +1 & +1 \end{bmatrix}. \quad (11.42)$$

Спектры планов можно изобразить в привычной для экспериментатора табличной форме. В табл. 11.1 приведен спектр плана ПФЭ 2^2 , а в табл. 11.2 — спектр плана ПФЭ 2^3 .

Таблица 11.1

i	Факторы	
	x_1	x_2
1	-1	-1
2	+1	-1
3	-1	+1
4	+1	+1

Таблица 11.2

i	Факторы			i	Факторы		
	x_1	x_2	x_3		x_1	x_2	x_3
1	-1	-1	-1	5	-1	-1	+1
2	+1	-1	-1	6	+1	-1	+1
3	-1	+1	-1	7	-1	+1	+1
4	+1	+1	-1	8	+1	+1	+1

В табл. 11.1, 11.2 и в последующем буквой i обозначен номер точки спектра плана.

Точки плана ПФЭ 2^n располагаются в вершинах n -мерного гиперкуба. На рис. 11.4, а показано расположение точек для двумерного случая, а на рис. 11.4, б — для трехмерного.

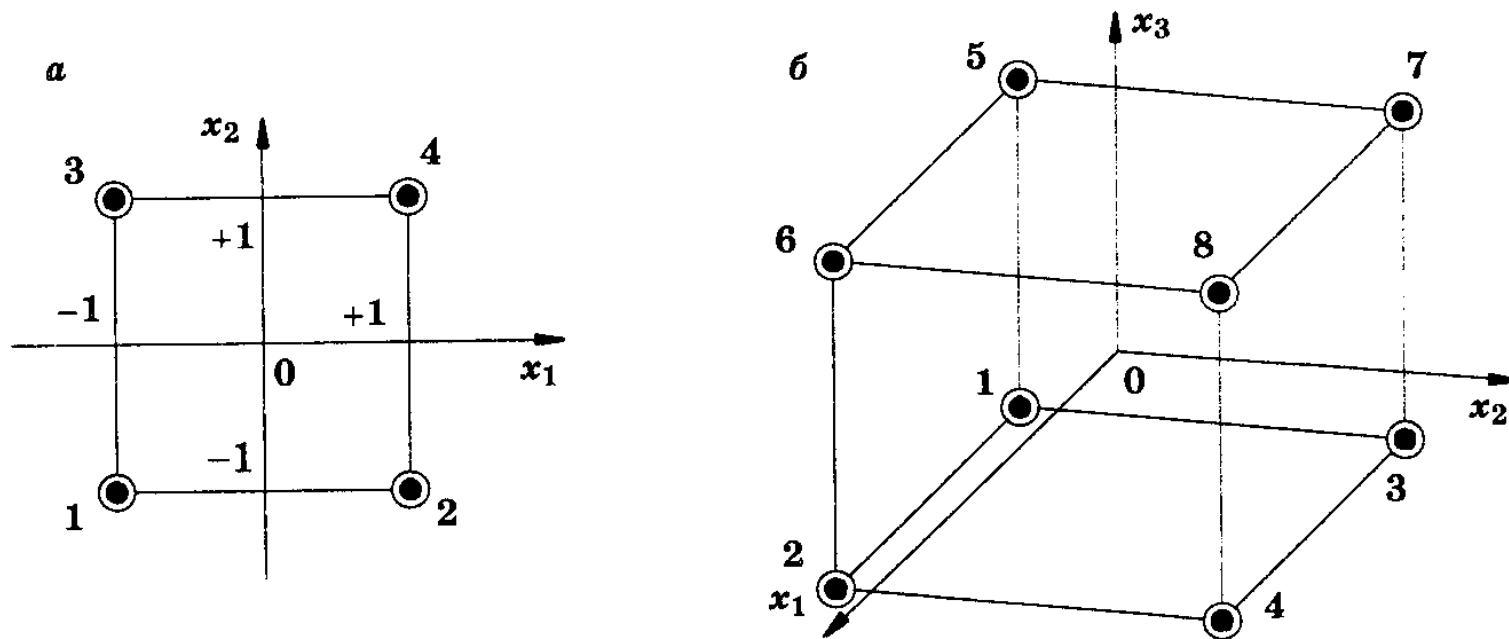


Рис. 11.4. Расположение точек спектра плана ПФЭ 2^n :
 а — при $n = 2$; б — при $n = 3$

Посредством ПФЭ можно построить как простейшую линейную модель технической системы вида

$$y = b_0 + b_1 x_1 + b_2 x_2 + \dots + b_n x_n, \quad (11.43)$$

так и нелинейную.

Для модели вида (11.43) система базисных функций очевидна: $f_0(\vec{X}) = 1$; $f_1(\vec{X}) = x_1$; $f_2(\vec{X}) = x_2$; ...; $f_n(\vec{X}) = x_n$. Число базисных функций в этом случае равно $n + 1$.

Выясним, какие базисные функции могут входить в регрессионную модель, получаемую посредством ПФЭ 2^n , чтобы выпол-

нялось требование о линейной независимости векторов-столбцов этих функций, изложенное в разделе 11.4. При выполнении этого требования получают отдельные оценки всех коэффициентов регрессии. Линейная независимость столбцов матрицы F достигается, если в ней отсутствуют полностью совпадающие или полностью противоположные (по знакам) столбцы.

В общем случае в полиномиальную модель могут входить факторы в любой степени и различные комбинации из их произведений. Так как при нормированных факторах их значения равны $+1$ или -1 , а в качестве показателей степеней факторов принимаются целые числа, то при четных показателях степеней вектор-столбец базисной функции состоит только из $+1$ и совпадает с вектором-столбцом функции $f_0(\bar{X})$, а векторы-столбцы всех базисных функций, соответствующих одним и тем же факторам x_j , возведенным в любые нечетные степени, будут совпадающими. Вместе с тем легко убедиться, что любые комбинации произведений факторов x_1, x_2, \dots, x_n могут быть в числе базисных функций.

Выпишем выражения линейных регрессий при $n = 2$ с учетом всех возможных сочетаний взаимодействия факторов

$$y = b_0 + b_1x_1 + b_2x_2 + b_3x_1x_2. \quad (11.44)$$

При $n = 3$ получаем

$$y = b_0 + b_1x_1 + b_2x_2 + b_3x_3 + b_4x_1x_2 + b_5x_1x_3 + b_6x_2x_3 + b_7x_1x_2x_3. \quad (11.45)$$

В табл. 11.3 приведены базисные функции плана ПФЭ 2^2 , используемого для построения регрессионной модели (11.44), а в табл. 11.4 — плана ПФЭ 2^3 , используемого для модели (11.45). Прямоугольниками в этих таблицах обведены спектры планов.

Таблица 11.3

i	$f_0 = 1$	$f_1 = x_1$	$f_2 = x_2$	$f_3 = x_1x_2$
1	+1	-1	-1	+1
2	+1	+1	-1	-1
3	+1	-1	+1	-1
4	+1	+1	+1	+1

Уравнение линейной регрессии, как это видно из (11.39) и (11.45), может содержать следующее предельное количество коэффициентов при различных видах базисных функций:

один коэффициент b_0 — свободный член уравнения регрессии;

n коэффициентов b_j — линейных членов уравнения регрессии.

C_n^2 коэффициентов $b_{j,k}$ при парных взаимодействиях факторов;

C_n^3 коэффициентов $b_{j,k,l}$ при тройных взаимодействиях факторов и т.д.;

один коэффициент $b_{1,2,\dots,n}$ при взаимодействии факторов максимального, n -го порядка.

Таблица 11.4

i	$f_0 = 1$	$f_1 = x_1$	$f_2 = x_2$	$f_3 = x_3$	$f_4 = x_1x_2$	$f_5 = x_1x_3$	$f_6 = x_2x_3$	$f_7 = x_1x_2x_3$
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

Выражение для определения общего числа коэффициентов регрессии имеет вид

$$N_B = 1 + n + C_n^2 + C_n^3 + \dots + C_n^n = 2^n. \quad (11.46)$$

Так как при использовании всех возможных сочетаний факторов в уравнении регрессии число определяемых коэффициентов N_B равно числу точек N спектра плана ПФЭ 2^n , то такой план является насыщенным.

Численные значения $f_j(\bar{X}_i)$, приведенные в таблице базисных функций, являются элементами матрицы F . Матрица F плана ПФЭ 2^n обладает следующими свойствами:

1. *Свойством симметричности относительно центра эксперимента* — алгебраическая сумма элементов каждого столбца матрицы базисных функций, кроме столбца $f_0(\bar{X})$, равна нулю:

$$\sum_{i=1}^N f_j(\bar{X}_i) = 0, \quad j = \overline{1, d}, \quad d = N_B - 1, \quad (11.47)$$

где $f_j(\bar{X}_i)$ — значение j -й базисной функции, соответствующее i -й строке матрицы F ; i — номер точки спектра плана; N — число точек спектра плана; N_B — количество базисных функций.

2. *Свойством ортогональности столбцов* — сумма построчных произведений элементов любых двух столбцов равна нулю:

$$\sum_{i=1}^N f_j(\bar{X}_i) f_k(\bar{X}_i) = 0, \quad j \neq k; \quad j, k = \overline{0, d}. \quad (11.48)$$

3. *Свойством нормировки* — сумма квадратов элементов каждого столбца матрицы базисных функций равна числу точек N спектра плана:

$$\sum_{i=1}^N [f_j(\bar{X}_i)]^2 = N, \quad j = \overline{0, d}. \quad (11.49)$$

4. Для столбца базисной функции $f_0(\bar{X})$ сумма элементов также равна N :

$$\sum_{i=1}^N f_0(\bar{X}_i) = N. \quad (11.50)$$

Выражения (11.47) — (11.50) записаны в предположении, что дублирование опытов не производится.

Составим информационную матрицу Фишера Φ , определяемую выражением (11.22). Выражения (11.48) и (11.49) позволяют определить элементы матрицы Φ . Очевидно, что для ПФЭ 2^n матрица Φ диагональная с постоянными диагональными элементами:

$$\Phi = \begin{bmatrix} N & 0 & \dots & 0 \\ 0 & N & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & N \end{bmatrix}. \quad (11.51)$$

Следовательно, ПФЭ 2^n относится к классу *ортогональных планов*.

Так как матрица Φ диагональная, то корреляционные моменты оценок коэффициентов регрессии $\mu_{11}(b_j, b_k) = 0$ и оценки всех коэффициентов регрессии $b_j, j = \overline{1, N_B}$, некоррелированы друг с другом. Кроме того, все коэффициенты регрессии оцениваются с одинаковой точностью, так как диагональные элементы матрицы Φ одинаковы. Дисперсия оценок коэффициентов

$$\sigma_{b_j}^2 = \sigma_\varepsilon^2 / N. \quad (11.52)$$

Для линейной модели вида (11.43) план ПФЭ 2^n является *A- и E-оптимальным* и *ротатабельным*, а для модели (11.39) — *D-оптимальным*.

ПЛАН ДРОБНОГО ФАКТОРНОГО ЭКСПЕРИМЕНТА

Наряду с отмеченными положительными качествами полного факторного эксперимента он имеет существенный недостаток: увеличение количества факторов приводит к быстрому росту числа опытов, что обусловлено степенной зависимостью (11.38). Например, при $n = 10$ спектр плана содержит $N = 2^{10} = 1024$ опыта. Кроме того, необходимо дублирование опытов.

ПФЭ позволяет построить регрессионную модель, которая учитывает влияние на функцию отклика выбранных факторов и всех возможных сочетаний взаимодействий этих факторов. Но поскольку структура модели выбирается на основе априорной информации о физических свойствах исследуемого объекта, то весьма сложно представить себе влияние на характеристики его функционирования эффектов взаимодействий выше второго или третьего порядка. Обычно при построении многофакторной регрессионной модели ограничиваются парными или, в крайнем случае, отдельными тройными взаимодействиями факторов. В этом случае ПФЭ оказывается избыточным, так как число точек спектра плана N значительно больше количества коэффициентов регрессии N_B . В результате возникает возможность сокращения числа опытов. Но при этом, естественно, должно соблюдаться условие возможности оценки коэффициентов регрессии по результатам опытов, которое выражается соотношением $N \geq N_B$.

Во многих случаях на начальной стадии моделирования технической системы в связи с отсутствием необходимой информации о влиянии на ее выходные параметры различных факторов (внутренних или внешних параметров) строят линейную модель вида (11.43). Например, при трех факторах выбирают модель в виде

$$y = b_0 + b_1x_1 + b_2x_2 + b_3x_3. \quad (11.53)$$

В этом уравнении четыре коэффициента регрессии, а при $n = 3$ спектр плана ПФЭ, согласно выражению (11.40), содержит 8 точек, т.е. предусматривает 8 опытов в различных точках факторного пространства. Следовательно, четыре опыта оказываются избыточными и их можно было бы исключить, естественно, при условии выполнения принятых предпосылок регрессионного анализа, прежде всего ортогональности столбцов матрицы базисных функций F .

При построении математических моделей, использующих упрощенные уравнения регрессий, когда $N > N_B$, применяют *дробные факторные эксперименты* (ДФЭ). Наибольшее

распространение имеют регулярные планы ДФЭ типа 2^{n-p} , т.е. ДФЭ 2^{n-p} , где n — число факторов, p — степень дробности ДФЭ. Планы ДФЭ принято называть *репликами* с указанием их степени дробности. Так, план ДФЭ 2^{n-1} называют полурепликой ПФЭ 2^n (1/2-реплика); ДФЭ 2^{n-2} — 1/4-реплика ПФЭ 2^n ; ДФЭ 2^{n-3} — 1/8-реплика ПФЭ 2^n и т.д. Полуреплика сокращает число опытов в два раза по сравнению с ПФЭ, 1/4-реплика — в четыре раза и т.д.

При построении матрицы спектра плана ДФЭ 2^{n-p} необходимо обеспечить выполнение условий, описываемых выражениями (11.47) — (11.50), принимая во внимание, что число точек спектра этого плана определяется по формуле

$$N = 2^{n-p}. \quad (11.54)$$

Условия (11.47) — (11.50) удовлетворяются, если в матрице базисных функций F отсутствуют полностью совпадающие или полностью противоположные столбцы, что позволяет получить раздельное оценивание всех коэффициентов регрессии.

При выборе степени дробности ДФЭ должно выполняться условие

$$N \geq N_B. \quad (11.55)$$

Выбранные базисные функции для ДФЭ составляют лишь некоторую часть базисных функций соответствующего ПФЭ. Назовем эти функции *существенными переменными*, характеризующими в наибольшей мере физические свойства технического объекта.

Процедура построения спектра плана ДФЭ 2^{n-p} содержит четыре этапа.

Этап 1. Выбор структуры уравнения регрессии и определение степени дробности ДФЭ. При этом исходят из условия выполнения соотношения (11.55).

Этап 2. Выбор ведущих факторов и построение для них матрицы спектра плана, определяющего программу их изменения в ходе эксперимента.

Число k ведущих факторов принимают равным разности между количеством факторов n и степенью дробности ДФЭ:

$$k = n - p. \quad (11.56)$$

Для выбранных ведущих факторов x_1, x_2, \dots, x_k строят план ПФЭ 2^k , используя изложенное в предыдущем параграфе правило чередования знаков.

Этап 3. Построение матрицы X спектра плана ДФЭ 2^{n-p} . Часть этой матрицы составляет матрица спектра плана ПФЭ 2^k , а

во вторую часть должны войти столбцы матрицы для остальных факторов $x_{k+1}, x_{k+2}, \dots, x_n$, количество которых равно

$$p = n - k. \quad (11.57)$$

Столбцы матрицы X , соответствующие этим факторам, определяют путем перемножения соответствующих столбцов ведущих факторов. Для этого используют генерирующие соотношения. *Генерирующим соотношением* называется алгебраическое выражение, устанавливающее связь между одним из факторов $x_{k+1}, x_{k+2}, \dots, x_n$ и произведением какой-либо комбинации ведущих факторов x_1, x_2, \dots, x_k .

Чтобы получаемые столбцы были ортогональными, для каждого из них задается отдельное генерирующее соотношение (количество этих соотношений равно p). Выбор генерирующих соотношений, вообще говоря, произволен. Однако в качестве генерирующих нельзя использовать те произведения ведущих факторов, которые входят в состав существенных переменных, так как в этом случае в матрице базисных функций F окажутся совпадающие столбцы: для одного из факторов $x_{k+1}, x_{k+2}, \dots, x_n$ и одного из взаимодействий факторов из числа существенных переменных.

Генерирующее соотношение имеет вид

$$x_{k+i} = x_j x_l x_m \dots, i = \overline{1, p}, \quad (11.58)$$

где x_{k+i} — фактор, не включенный в число ведущих (для него определяется столбец матрицы X спектра плана ДФЭ 2^{n-p});
 x_j, x_l, x_m, \dots — ведущие факторы.

Количество ведущих факторов, входящих в генерирующее соотношение (11.58), может быть произвольным, но соотношения (11.58) для всех x_{k+i} должны быть разными.

Этап 4. Проверка пригодности полученного спектра плана.

Для этого необходимо построить матрицу базисных функций F и проверить, нет ли в ней совпадающих или полностью противоположных столбцов, т.е. выяснить, обладает ли матрица F свойством ортогональности столбцов, определяемым выражением (11.48). Если в матрице F нет совпадающих или противоположных столбцов, полученный спектр плана ДФЭ 2^{n-p} пригоден для решения поставленной задачи. В противном случае выполняются последовательно следующие процедуры до тех пор, пока не будет обеспечена ортогональность:

- выбираются иные генерирующие соотношения;
- изменяется набор ведущих факторов;
- уменьшается степень дробности плана p .

При ограниченных возможностях проведения опытов степень дробности плана сохраняют, а изменяют структуру уравнения регрессии (например, используют иные взаимодействия факторов или исключают какую-либо базисную функцию, соответствующую одному из взаимодействий высшего порядка).

Таким образом, регулярные планы $ДФЭ2^{n-p}$ обладают теми же свойствами, что и планы $ПФЭ2^n$. Матрица F удовлетворяет выражениям (11.47) — (11.50). Информационная матрица Фишера Φ диагональная и имеет вид (11.51). Дисперсию оценок коэффициентов регрессии определяют по формуле (11.52). Планы $ДФЭ2^{n-p}$ ортогональны. Для линейных моделей они ротатабельны, A - и E -оптимальны, а насыщенные планы D -оптимальны. Поскольку планы $ДФЭ$ значительно экономичнее планов $ПФЭ$, они получили широкое практическое применение. В частности, их используют для анализа чувствительности целевой функции к вариации параметров технических объектов в процессе их отсеивания и отбора для осуществления оптимизации.

Рассмотрим примеры построения планов ДФЭ.

Пример 11.1. Получить спектр плана ДФЭ, предназначенного для оценки коэффициентов уравнения регрессии вида (11.53).

Так как число факторов в этом уравнении $n = 3$, то при проведении ПФЭ количество точек спектра плана было бы равно $N = 2^3 = 8$. В уравнении же (11.53) всего четыре коэффициента, поэтому можно использовать полуреплику, т.е. ДФЭ 2^{3-1} , спектр плана которой содержит четыре точки: $N = 2^{3-1} = 4$, и следовательно, условие (11.55) выполняется.

Число ведущих факторов $k = n - p = 3 - 1 = 2$. Выберем в качестве ведущих факторов x_1 и x_2 . Значения элементов векторов-столбцов этих факторов получим на основе плана ПФЭ 2^2 , используя метод чередования знаков. Для определения вектора-столбца фактора x_3 примем генерирующее соотношение в виде $x_3 = x_1x_2$. Полученный спектр плана ДФЭ 2^{3-1} выделен прямоугольником в табл. 11.5, в которой приведена матрица базисных функций F .

Таблица 11.5

i	$f_0 = 1$	$f_1 = x_1$	$f_2 = x_2$	$f_3 = x_3$
1	+1	-1	-1	+1
2	+1	+1	-1	-1
3	+1	-1	+1	-1
4	+1	+1	+1	+1

В матрице F нет совпадающих столбцов, следовательно, полученный спектр плана пригоден для решения поставленной задачи.

Пример 11.2. Задан список существенных переменных: $x_1, x_2, x_3, x_4, x_1x_2, x_2x_3, x_3x_4$. Получить спектр плана ДФЭ.

Уравнение регрессии в этом случае будет включать 4 фактора и 8 коэффициентов, а число точек спектра плана ПФЭ равно $N = 2^4 = 16$. Следовательно, можно попытаться использовать ДФЭ 2^{4-1} , спектр плана которого содержит необходимое число точек $N = 8$ и обеспечивает выполнение условия (11.55).

Векторы-столбцы x_1x_2 и x_3x_4 оказались одинаковыми, следовательно, полученный план непригоден. Последовательно перебирая все возможные варианты решения проблемы, можно убедиться, что ни один из них не дает положительного результата. Это означает, что при заданном списке существенных переменных план ДФЭ2⁴⁻¹ не может быть применен для получения искомого уравнения регрессии. Следовательно, необходимо использовать план ПФЭ2⁴.

Пример 11.3. Введем небольшое изменение в список существенных переменных примера 11.2: вместо x_3x_4 примем x_2x_4 .

Те же действия, что и в предыдущем примере, дают значения элементов матрицы F , приведенные в табл. 11.7.

Таблица 11.7

i	$f_0 = 1$	$f_1 = x_1$	$f_2 = x_2$	$f_3 = x_3$	$f_4 = x_4$	$f_5 = x_1x_2$	$f_6 = x_2x_3$	$f_7 = x_2x_4$
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

В матрице F все столбцы различны и она пригодна для получения уравнения регрессии выбранной структуры.

Следует отметить, что генерирующие соотношения можно принимать также с обратным знаком, например $x_4 = -x_1x_2x_3$. В этом случае лишь поменяются

знаки в векторах-столбцах базисных функций $f_4 = x_4$ и $f_7 = x_2x_4$ на противоположные, но это не приведет к изменению свойств матрицы F , оцениваемых выражениями (11.47) – (11.50).

Если принять генерирующее соотношение $x_4 = x_1x_3$, то спектр плана ДФЭ 2^{4-1} также окажется пригодным, что легко проверить.

**ГЕНЕРИРУЮЩИЕ
СООТНОШЕНИЯ И
ОПРЕДЕЛЯЮЩИЕ КОНТРАСТЫ**

Если матрица базисных функций F плана ДФЭ содержит полностью совпадающие или противоположные столбцы, то это исключает возможность отдельного оценивания коэффициентов регрессии, соответствующих этим столбцам. В этом случае говорят, что имеет место смешанная оценка. Коэффициенты регрессии при этом оказываются коррелированными.

Для выяснения системы смешивания используют определяющие контрасты. *Определяющие контрасты* матрицы F представляют собой произведения левых и правых частей генерирующих соотношений.

Генерирующее соотношение, устанавливающее связь между фактором x_{k+i} и ведущими факторами x_j, x_l, x_m , имеет вид

$$x_{k+i} = x_j x_l x_m. \quad (11.59)$$

Умножим левую и правую части этого выражения на x_{k+i} . Учитывая, что элементы столбца матрицы F при нормированных факторах равны $+1$ или -1 , получим $x_{k+i}x_{k+i} = 1$. В результате определяющий контраст будет представлен выражением

$$1 = x_j x_l x_m x_{k+i}. \quad (11.60)$$

Число определяющих контрастов, так же как и число генерирующих соотношений ДФЭ, равно p . Составив выражения определяющих контрастов для всех генерирующих соотношений и перемножив их левые и правые части, получим *обобщенный определяющий контраст* (ООК). При этом в правой части полученного произведения может оказаться некоторое количество одинаковых сомножителей, например t сомножителей фактора x_j . В этом случае следует иметь в виду, что при четном числе сомножителей результат их перемножения равен единице, а при нечетном равен x_j , т.е.

$$\prod_1^m x_j = 1 \quad \text{при } m = 2, 4, 6, \dots,$$

$$\prod_1^m x_j = x_j \quad \text{при } m = 3, 5, 7, \dots$$

Для определения системы смешивания влияния переменных в ДФЭ обобщенный определяющий контраст умножают на каждую переменную, т.е. на каждую базисную функцию. Если при этом получают одинаковые результаты для какой-либо пары базисных функций, то соответствующие столбцы матрицы F окажутся совпадающими и оценки коэффициентов регрессии при этих базисных функциях будут коррелированными (смешанными). Такой план ДФЭ непригоден для получения регрессии заданной структуры. В общем случае может оказаться несколько пар совпадающих столбцов матрицы F .

В табл. 11.8 приведены результаты определения системы смешивания влияния переменных в ДФЭ 2^{4-1} для примера 11.2. В этом примере использовано одно генерирующее соотношение: $x_4 = x_1 x_2 x_3$. В результате выражение ООК имеет вид: $1 = x_1 x_2 x_3 x_4$.

Таблица 11.8

Переменные уравнения регрессии	ООК	Система смешивания
x_0		$x_0 = x_1 x_2 x_3 x_4$
x_1		$x_1 = x_2 x_3 x_4$
x_2		$x_2 = x_1 x_3 x_4$
x_3	$1 = x_1 x_2 x_3 x_4$	$x_3 = x_1 x_2 x_4$
x_4		$x_4 = x_1 x_2 x_3$
$x_1 x_2$		<u>$x_1 x_2 = x_3 x_4$</u>
$x_2 x_3$		$x_2 x_3 = x_1 x_4$
$x_3 x_4$		<u>$x_3 x_4 = x_1 x_2$</u>

Одинаковые результаты оценок системы смешивания подчеркнуты. Из таблицы видна смешанность оценок влияния переменных $x_1 x_2$ и $x_3 x_4$ на функцию отклика, что не дает возможности осуществить отдельную оценку коэффициентов регрессии при этих переменных.

**СТАТИСТИЧЕСКИЙ АНАЛИЗ
РЕЗУЛЬТАТОВ АКТИВНОГО
ЭКСПЕРИМЕНТА**

Прежде чем определять коэффициенты регрессии, необходимо выполнить статистический анализ результатов эксперимента с целью оценки их качества и пригодности для построения регрессионной модели. Статистический анализ включает оценку ошибок параллельных опытов, отсеивание грубых ошибок, проверку однородности дисперсий опытов и определение дисперсии воспроизводимости эксперимента.

Ошибки параллельных опытов. В условиях наличия случайных помех с целью уменьшения случайных погрешностей эксперимента и повышения точности получаемой регрессионной модели осуществляется дублирование опытов, т.е. проведение параллельных опытов. Каждый опыт, предусмотренный матрицей спектра плана, повторяется $m = 2...5$ раз. Рекомендуется число m принимать одинаковым для всех N точек плана. В результате проводится $L = Nm$ опытов в соответствии с матрицей плана, предусматривающей при этом рандомизацию опытов.

Повторные опыты в одной и той же точке плана при наличии помехи дают различные результаты при определении функции отклика. Разброс результатов относительно оценки математического ожидания функции отклика называют *ошибкой воспроизводимости опыта*. Эту ошибку надо оценить.

Для каждой точки плана по результатам параллельных опытов находят *выборочное среднее* \bar{y}_i , равное среднему арифметическому полученных опытных значений функции отклика

$$\bar{y}_i = \frac{1}{m} \sum_{u=1}^m y_{iu}, \quad i = \overline{1, N}, \quad (11.61)$$

где u — номер параллельного опыта; y_{iu} — значение функции отклика в u -м параллельном опыте i -й точки спектра плана.

Для оценки отклонения функции отклика от ее среднего значения \bar{y}_i вычисляется *дисперсия воспроизводимости* опыта по данным m параллельных опытов в каждой i -й точке спектра плана

$$S_i^2 = \frac{1}{m-1} \sum_{u=1}^m (y_{iu} - \bar{y}_i)^2, \quad i = \overline{1, N}. \quad (11.62)$$

При вычислении S_i^2 принимают число степеней свободы k на единицу меньше, чем число параллельных опытов, т.е. $k = m - 1$, так как одна степень свободы уже использована для

вычисления \bar{y}_i . Это обеспечивает несмещенность оценки дисперсии воспроизводимости опыта S_i^2 .

Отсеивание грубых ошибок. Формула (11.61) справедлива лишь при нормальном распределении случайной величины y . При наличии грубых ошибок опыта распределение y отклоняется от нормального, что противоречит предпосылкам 1 и 2 (см. раздел 11.4), положенным в основу регрессионного анализа. Поэтому грубые ошибки надо вначале исключить, а затем определять \bar{y}_i и S_i^2 . Грубые ошибки — это брак повторных опытов. Для обнаружения брака используют t -критерий Стьюдента

$$t_{iu} = (y_{iu} - \bar{y}_i^*) / S_i^*, \quad (11.63)$$

где S_i^* — среднее квадратическое отклонение.

Значения \bar{y}_i^* и S_i^* определяются по формулам (11.61) и (11.62), но без учета оцениваемого результата опыта y_{iu} .

Полученное значение t -критерия сравнивается с табличным t_T при выбранном уровне значимости $q = P[t > t_{k,q}]$ и числе степеней свободы k . Уровень значимости q характеризует вероятность ошибки. Если $t > t_T$, то это соответствует браку данного опыта и результат его не может быть использован. В этом случае опыт подлежит повторному проведению.

Значения t -критерия Стьюдента в зависимости от числа степеней свободы k и уровня значимости q приведены в табл. 4 приложения.

Проверка однородности дисперсий. Принимается нулевая гипотеза об однородности дисперсий воспроизводимости опытов. Однородность дисперсий означает, что среди всех дисперсий S_i^2 нет таких, которые бы значительно превышали все остальные. Для проверки однородности дисперсий во всех точках спектра плана используется либо критерий Кохрена G , либо критерий Фишера F (табл. 6 и 5 приложения).

Критерий Кохрена основан на распределении отношения максимальной дисперсии $S_{i \max}^2$ к сумме всех дисперсий:

$$G = S_{i \max}^2 / \sum_{i=1}^N S_i^2. \quad (11.64)$$

Критерий Кохрена применяется, если количество сравниваемых дисперсий больше двух, а число повторных опытов во всех точках плана одинаково. Определив число степеней свободы $k_1 = m - 1$ и $k_2 = N$ (N — число точек спектра плана, m — количество повторных опытов в каждой точке плана), находят табличное значение критерия Кохрена G_T . Если $G < G_T$, гипотеза об однородности дисперсий и воспроизводимости результатов принимается. Это означает, что предпосылки 1 и 2, положенные в основу регрессионного анализа, выполняются. В этом случае каждая из дисперсий S_i^2 оценивает одну и ту же дисперсию помехи σ_ε^2 . Следовательно, полученные результаты эксперимента качественные и могут быть использованы для построения регрессионной модели. В противном случае следует увеличить число параллельных опытов или повторить эксперимент при строгом соблюдении методики и схемы проведения опытов, предприняв необходимые меры для исключения грубых ошибок.

Если выяснится, что непостоянство дисперсии помехи σ_{ε}^2 обусловлено внутренними свойствами объекта, то необходимы более сложные способы обработки результатов эксперимента. Можно, например, вводить некоторую функцию от y : $\ln y$, \sqrt{y} и др.

Критерий Фишера позволяет сравнивать две дисперсии и определяется из соотношения

$$F = S_{\max}^2 / S_{\min}^2. \quad (11.65)$$

Дисперсии однородны, если $F < F_T$, где F_T — табличное значение критерия Фишера, определяемое при числах степеней свободы k_1 и k_2 и принятом уровне значимости q .

Следует отметить, что уровень значимости q по всем критериям, применяемым в процессе статистического анализа и обработки результатов эксперимента (Кохрена, Стьюдента, Фишера), должен быть одинаков. Для технических систем рекомендуется принимать $q = 0,05$.

Дисперсия воспроизводимости эксперимента. Если дисперсии S_i^2 однородны, то их усредняют и находят *дисперсию воспроизводимости эксперимента*

$$S_y^2 = \frac{1}{N} \sum_{i=1}^N S_i^2. \quad (11.66)$$

Дисперсия S_y^2 представляет собой оценку дисперсии помехи σ_ε^2 . Так как число степеней свободы при определении дисперсии S_i^2 равно $k = m - 1$, то число степеней свободы, связанное с оценкой S_y^2 , вычисляется по формуле

$$k = N(m - 1). \quad (11.67)$$

Формула (11.66) годится, если число повторных опытов во всех точках спектра плана одинаково. Если число опытов различно, используют формулу

$$S_y^2 = \frac{S_1^2 k_1 + S_2^2 k_2 + \dots + S_N^2 k_N}{k_1 + k_2 + \dots + k_N} = \left(\sum_{i=1}^N S_i^2 k_i \right) / \left(\sum_{i=1}^N k_i \right), \quad (11.68)$$

где k_i — число степеней свободы в i -й точке спектра плана; $k_i = m_i - 1$; m_i — число параллельных опытов в этой точке.

**ОПРЕДЕЛЕНИЕ КОЭФФИЦИЕНТОВ
РЕГРЕССИОННОЙ МОДЕЛИ ПРОВЕРКА ИХ
ЗНАЧИМОСТИ**

Параметрами регрессионной модели являются коэффициенты регрессии b_j , $j = \overline{0, N_B - 1}$, где N_B — количество базисных функций. Значения коэффициентов регрессии можно получить, решив систему алгебраических уравнений (11.20). В этих уравнениях величина индекса d в обозначении базисных функций f_{id} и коэффициента регрессии b_d равна $d = N_B - 1$. Так как информационная матрица Фишера Φ для ПФЭ и ДФЭ диагональная (11.51) и все диагональные элементы ее одинаковы и равны N , то выражение для определения всех коэффициентов уравнения регрессии одинаково и имеет простой вид:

$$b_j = \frac{1}{N} \sum_{i=1}^N f_j(\bar{X}_i) \bar{y}_i, \quad (11.69)$$

где N — число точек спектра плана; $f_j(\bar{X}_i)$ — значение j -й базисной функции в i -й точке спектра плана; \bar{y}_i — выборочное среднее функции отклика в той же точке, определяемое по формуле (11.61).

Значения базисных функций $f_j(\bar{X}_i)$ для отдельных факторов равны X_{ij} , а для взаимодействия факторов — $X_{ik}X_{il}X_{im}\dots$. С учетом этого на основе выражения (11.69) можно записать сле-

дующие формулы для вычисления значений коэффициентов уравнения регрессии:

для коэффициентов при факторах x_j , включая также свободный член уравнения,

$$b_j = \frac{1}{N} \sum_{i=1}^N X_{ij} \bar{y}_i, \quad j = \overline{0, n}; \quad (11.70)$$

для коэффициентов при взаимодействиях факторов

$$b_g = \frac{1}{N} \sum_{i=1}^N X_{ik} X_{il} X_{im} \dots \bar{y}_i, \quad g = \overline{n+1, d}; \quad (11.71)$$

$$k, l, m = \overline{1, n}; \quad k \neq l \neq m,$$

где n — количество факторов.

Формулы (11.70) и (11.71) применяются для планов первого порядка. Для плана ПФЭ $N = 2^n$, а для ДФЭ $N = 2^{n-p}$. При определении коэффициента b_0 (свободного члена уравнения регрессии) $X_{i0} = 1, i = \overline{1, N}$.

Поскольку полученные значения коэффициентов регрессии $b_j, j = \overline{0, N_B - 1}$, — случайные числа, в связи с действием случайной помехи в процессе эксперимента, то они являются оценками истинных значений коэффициентов регрессии β_j . Погрешность

определения b_j оценивают дисперсией $S_{b_j}^2$.

Дисперсии $S_{b_j}^2$ оценок всех коэффициентов регрессии, как показано в разделах 11.8 и 11.9, одинаковы. Величина дисперсии $S_{b_j}^2$ зависит только от ошибки воспроизводимости эксперимента S_y^2 и числа опытов:

$$S_{b_j}^2 = S_y^2 / (Nm), \quad (11.72)$$

где m — число повторных опытов (значение m должно быть одинаковым для всех точек N спектра плана).

После определения коэффициентов регрессии b_j проверяют их значимость. Принимается нулевая гипотеза о незначимости полученных коэффициентов и отсутствии влияния соответствующих им базисных функций на функцию отклика y . Проверка гипотезы осуществляется с использованием t -критерия Стьюдента, значение которого находят из соотношения

$$t_j = |b_j| / S_{b_j}, j = \overline{0, N_B - 1}, \quad (11.73)$$

где S_{b_j} — среднее квадратическое отклонение погрешности оценки b_j ; N_B — общее число коэффициентов уравнения регрессии, равное количеству используемых базисных функций для построения регрессии.

Полученное значение t_j для каждого коэффициента регрессии b_j сравнивают с табличным t_T , определяемым при принятом уровне значимости q и числе степеней свободы $k = N(m - 1)$, с которым определялась дисперсия воспроизводимости S_y^2 . Если $t_j < t_T$, нулевая гипотеза о незначимости коэффициента b_j принимается и член уравнения регрессии, включающий этот коэффициент, исключается из математической модели. Если же $t_j > t_T$, полагают, что данный коэффициент значимо (неслучайно) отличается от нуля и его следует сохранить в регрессионной модели. В этом случае значение коэффициента b_j больше ошибки опыта, которую можно оценить величиной доверительного интервала ε_{b_j} . Доверительный интервал находят по формуле

$$\varepsilon_{b_j} = \pm t_T S_{b_j}. \quad (11.74)$$

Следует, однако, отметить, что дисперсия воспроизводимости эксперимента S_y^2 зависит от очень многих факторов: выбора центра эксперимента, интервалов варьирования факторов, наличия экстремумов функции отклика в области планирования, соотношения величины отклика и помехи (так называемое отношение сигнал — шум) и др. В этой связи при небольшом различии между t_j и t_T следует весьма осторожно относиться к оценке значимости коэффициентов регрессии. Лучше такие коэффициенты сохранить в модели, а влияние соответствующего фактора (или взаимодействия факторов) проверить в дальнейшем на более сложной модели или в иных условиях планирования эксперимента.

После исключения незначимых коэффициентов уравнение регрессии приобретает вид

$$\hat{y} = \sum_{j=0}^{N_B^* - 1} b_j f_j(\bar{X}), \quad (11.75)$$

где N_B^* — количество значимых коэффициентов регрессии.

Так как часть коэффициентов регрессии исключена из модели, то $N_{\text{в}}^* < N_{\text{в}} \leq N$.

Если все коэффициенты оказались значимыми, суммирование в формуле (11.75) осуществляется до $N_{\text{в}} - 1$.

**ПРОВЕРКА АДЕКВАТНОСТИ И
РАБОТОСПОСОБНОСТИ
РЕГРЕССИОННОЙ МОДЕЛИ**

По уравнению регрессии (11.75) можно вычислить предсказанные значения функции отклика \bar{y} во всех точках спектра плана: $\bar{X}_1, \bar{X}_2, \dots, \bar{X}_N$. В результате будет получено N значений $\hat{y} : \hat{y}_1, \hat{y}_2, \dots, \hat{y}_N$. Если регрессионная модель получена на основе ПФЭ и все коэффициенты регрессии признаны значимыми, то в формуле (11.75) $N_{\text{в}}^* = N_{\text{в}} = N$. Тогда значения \hat{y}_i должны совпадать со средними выборочными значениями \bar{y}_i , полученными в результате эксперимента для каждой точки спектра плана. Следовательно, поверхность отклика $Y = \varphi(\bar{X})$ проходит через все точки $\bar{y}_i, i = \overline{1, N}$, и полученная модель адекватна. Значения \hat{y}_i в этом случае используют для проверки правильности вычислений коэффициентов регрессии.

Если же незначимые коэффициенты b_j исключены из регрессионной модели, то $N_B^* < N$. Тогда $\hat{y}_i \neq \bar{y}_i$. Это же характерно для моделей, полученных на основе ДФЭ. Разности $(\hat{y}_i - \bar{y}_i)$ несут информацию об ошибках предсказания по уравнению регрессии и их можно использовать для последующего анализа свойств полученной модели — ее адекватности и работоспособности.

Для оценки рассеяния эмпирических значений \bar{y}_i относительно расчетных \hat{y}_i , полученных по уравнению регрессии, используют дисперсию адекватности

$$S_{\text{ад}}^2 = \frac{m}{N - N_B^*} \sum_{i=1}^N (\bar{y}_i - \hat{y}_i)^2, \quad (11.76)$$

где m — число параллельных опытов; N — число точек спектра плана; N_B^* — количество значимых коэффициентов регрессии.

Если число параллельных опытов в различных точках спектра плана неодинаково, то для вычисления $S_{\text{ад}}^2$ используют формулу

$$S_{\text{ад}}^2 = \frac{1}{N - N_{\text{в}}^*} \sum_{i=1}^N m_i (\bar{y}_i - \hat{y}_i)^2, \quad (11.77)$$

где m_i — число параллельных опытов в i -й точке спектра плана.

При оценке регрессионной модели принимается нулевая гипотеза о том, что полученная модель обеспечивает адекватное описание результатов эксперимента. Проверка адекватности осуществляется путем сопоставления дисперсии адекватности $S_{\text{ад}}^2$ и дис-

персии воспроизводимости эксперимента S_y^2 . У адекватной моде-

ли значение $S_{\text{ад}}^2$ обусловлено в основном действием случайной

помехи, поэтому различие между $S_{\text{ад}}^2$ и S_y^2 должно быть неболь-

шим, так как они оценивают одну и ту же дисперсию помехи σ_{ε}^2 .

Проверку гипотезы об адекватности модели (гипотезы о равенстве дисперсий $S_{\text{ад}}^2$ и S_y^2) выполняют по критерию Фишера

$$F = S_{\text{ад}}^2 / S_y^2. \quad (11.78)$$

В формулах (11.76) и (11.77) учтено, что чем больше число m параллельных опытов, тем с большей достоверностью оцениваются средние значения функции отклика y . Поэтому требования к различиям между экспериментальными \bar{y}_i и расчетными \hat{y}_i значениями становятся более жесткими, что отражается в увеличении F -критерия.

Полученные значения статистики F сравнивают с табличным значением критерия Фишера F_T , определяемым в зависимости от уровня значимости q и чисел степеней свободы k_1 и k_2 , с которыми определялись дисперсии $S_{ад}^2$ и S_y^2 :

$$k_1 = N - N_B^*; \quad (11.79)$$

$$k_2 = N(m - 1). \quad (11.80)$$

Если $F < F_T$, регрессионная модель считается адекватной.

Различие между дисперсиями $S_{ад}^2$ и S_y^2 обусловлено систематической ошибкой при определении функции отклика по уравнению регрессии из-за его приближенности. Если модель описывает физические свойства исследуемого объекта неудовлетворительно, систематическая ошибка приводит к значительному воз-

растанию дисперсии адекватности и, следовательно, к увеличению статистики F .

При $F > F_T$ гипотеза адекватности модели отвергается. В таком случае нужно либо изменить структуру математической модели, либо уменьшить интервалы варьирования факторов и провести повторно эксперимент с моделью прежней структуры.

В первом варианте реализуется принцип постепенного усложнения структуры математической модели. Если использовалось упрощенное уравнение регрессии первого порядка, учитывающее влияние на функцию отклика только факторов, или факторов и некоторого количества эффектов их взаимодействий низших порядков, что характерно для ДФЭ, то в модель можно дополнительно ввести новые члены, содержащие другие эффекты взаимодействия тех же порядков или более высоких порядков. Однако во многих случаях такой путь оказывается неэффективным, так как, согласно выражению (11.76), при увеличении количества членов уравнения регрессии и неизменном числе точек спектра плана N дисперсия адекватности может возрасти, несмотря на снижение разности $(\bar{y}_i - \hat{y}_i)$, поскольку при этом увеличива-

ется N_B^* и, следовательно, уменьшается знаменатель выражения (11.76). Кроме того, следует иметь в виду, что с увеличением порядка эффекта взаимодействия возрастает вероятность незначимости коэффициента регрессии b_j при этом эффекте. В этой связи наиболее целесообразно перейти к планированию второго порядка, используя регрессионное уравнение в виде полного квадратного полинома (см. раздел 11.14).

После обеспечения адекватности регрессионной модели осуществляют проверку ее работоспособности.

Адекватность регрессионной модели еще не гарантирует ее пригодность к практическому использованию в задачах прогнозирования и поиска оптимальных решений. Модель может оказаться неработоспособной из-за низкой ее точности. Для проверки работоспособности модели используют *коэффициент детерминации*, представляющий собой числовую интегральную характеристику точности уравнения регрессии. Его значение вычисляют по формуле

$$R^2 = 1 - \frac{(N - N_B^*)S_{ад}^2 + N(m - 1)S_y^2}{m \sum_{i=1}^N (\bar{y}_i - \bar{y})^2 + N(m - 1)S_y^2}, \quad (11.81)$$

где \bar{y} — среднее значение отклика:

$$\bar{y} = \frac{1}{N} \sum_{i=1}^N \bar{y}_i. \quad (11.82)$$

Модель считается работоспособной при $R^2 \geq 0,75$. В этом случае обеспечивается уменьшение ошибки предсказания, полученного по уравнению регрессии, по крайней мере, в 2 раза в сравнении с предсказанием по среднему значению отклика \bar{y} без учета влияния факторов \bar{X} на функцию отклика y .

ПЛАНЫ ВТОРОГО ПОРЯДКА

Планы второго порядка предназначены для получения регрессионной модели в виде полного квадратного полинома — полинома второй степени. Такой полином содержит основные эффекты, все парные взаимодействия и квадратичные эффекты

$$y = b_0 + \sum_{j=1}^n b_j x_j + \sum_{j=1}^n \sum_{k=j+1}^n b_{j,k} x_j x_k + \sum_{j=1}^n b_{jj} x_j^2. \quad (11.83)$$

Число коэффициентов уравнения регрессии в этом случае

$$N_B = 1 + 2n + C_n^2 = \left[(n + 1)(n + 2) \right] / 2, \quad (11.84)$$

что в $(n + 2) / 2$ раз больше, чем в линейной модели вида (11.43).

Соответственно возрастает и минимально необходимое число точек в спектре плана. Для получения квадратичной модели варьирование факторов в эксперименте должно осуществляться, по крайней мере, на трех уровнях.

Существует большое множество различных планов второго порядка. Систематизированное их изложение дается в специальной литературе [9]. Рассмотрим кратко лишь композиционные планы типа V_n , получившие широкое применение благодаря их экономичности и простой структуре. Эти планы содержат ядро ПФЭ 2^n илиДФЭ 2^{n-p} и включают $2n$ звездных точек, которые расположены на координатных осях на расстоянии $\pm\alpha$ от центра эксперимента. Величина α выбирается из условия минимизации обобщенной дисперсии оценок коэффициентов регрессии, что обеспечивает минимум объема эллипсоида рассеяния этих оценок. Следовательно, планы V_n построены с учетом критерия D -оптимальности. Величина α для этих планов оказывается равной 1 для всех n факторов, а область планирования представляет собой гиперкуб. Центральной точки планы V_n не содержат.

Если ядром плана является ПФЭ 2^n , то число точек спектра плана типа V_n определяют по формуле

$$N = 2^n + 2n, \quad (11.85)$$

а если ядро составляет планДФЭ 2^{n-p} , то

$$N = 2^{n-p} + 2n. \quad (11.86)$$

В качестве примера в табл. 11.9 приведена матрица спектра плана типа V_n при $n = 3$, ядром которого является план ПФЭ 2^3 .

Таблица 11.9

i	x_1	x_2	x_3	i	x_1	x_2	x_3
1	-1	-1	-1	9	-1	0	0
2	+1	-1	-1	10	+1	0	0
3	-1	+1	-1	11	0	-1	0
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				

Из приведенной таблицы легко видеть процедуру построения плана. Первые восемь точек составляют ядро плана V_n и соответствуют спектру плана ПФЭ 2^3 , а остальные шесть — звездные точки. В этих точках варьируется только один какой-либо фактор x_j на нижнем или верхнем уровне, а остальные находятся в центре эксперимента и их нормированные значения равны нулю.

Оценки коэффициентов регрессии вычисляются по формулам:

$$b_0 = \frac{1}{2(n-1)} \left(\sum_{i=N_1+1}^N \bar{y}_i - \frac{1}{2^{n-p-1}} \sum_{i=1}^{N_1} \bar{y}_i \right); \quad (11.87)$$

$$b_{jj} = \frac{1}{2} \sum_{i=N_1+1}^N X_{ij}^2 \bar{y}_i - b_0; \quad (11.88)$$

$$b_j = \frac{1}{2 + 2^{n-p}} \sum_{i=1}^N X_{ij} \bar{y}_i; \quad (11.89)$$

$$b_{jk} = \frac{1}{2^{n-p}} \sum_{i=1}^N X_{ij} X_{ik} \bar{y}_i. \quad (11.90)$$

В формулах (11.87) и (11.88) N_1 — число точек ядра спектра плана: $N_1 = 2^{n-p}$.

Если отдельные коэффициенты b_j и b_{jk} окажутся незначимыми, то их можно исключить из уравнения регрессии без пересчета остальных коэффициентов.

Дисперсии оценок коэффициентов регрессии различны, так как план V_n не обладает свойством ортогональности, а коэффициенты b_0 и b_{jj} коррелированы, поэтому в случае незначимости некоторых коэффициентов b_{jj} при исключении их из модели требуется уточнение оставшихся коэффициентов b_0 и b_{jj} .

**РЕГРЕССИОННЫЙ АНАЛИЗ РЕЗУЛЬТАТОВ
ВЫЧИСЛИТЕЛЬНОГО ЭКСПЕРИМЕНТА НА
ДЕТЕРМИНИРОВАННОЙ ТЕОРЕТИЧЕСКОЙ
МОДЕЛИ**

Детерминированная математическая модель характеризуется взаимно однозначным соответствием между внешним воздействием на моделируемую систему и ее реакцией на это воздействие. Поэтому в каждой точке спектра плана проводят только один опыт. План активного вычислительного эксперимента составляется в зависимости от вида регрессионной модели так же, как и для вероятностных математических моделей. При построении экспериментальных факторных моделей, предназначенных для решения задач оптимизации параметров технических объектов в процессе их функционального проектирования, используют планы первого и второго порядков.

Регрессионный анализ при экспериментах на детерминированных и вероятностных моделях включает одни и те же этапы: статистический анализ результатов эксперимента, получение коэффициентов регрессии b_j , оценка адекватности и работоспособности экспериментальной факторной модели. Однако содержания первого и третьего этапов в обоих случаях различны.

На первом этапе осуществляется построение модели среднего и ее статистический анализ. При этом определяют среднее значение функции отклика \bar{y} и дисперсию модели среднего S_y^2 , характеризующую рассеяние результатов эксперимента относительно \bar{y} и оценивающую погрешность модели среднего:

$$\bar{y} = \frac{1}{N} \sum_{i=1}^N y_i; \quad (11.91)$$

$$S_y^2 = \frac{1}{N-1} \sum_{i=1}^N (y_i - \bar{y})^2, \quad (11.92)$$

где y_i — значение функции отклика в i -й точке спектра плана; N — количество проведенных опытов, равное числу точек спектра плана.

Коэффициенты регрессии первого порядка определяются по формулам (11.70) и (11.71), а регрессии второго порядка, полученной на основе плана типа B_n , — по формулам (11.87) — (11.90).

После определения коэффициентов b_j осуществляется проверка пригодности полученного уравнения регрессии. Для этого вначале необходимо вычислить по уравнению регрессии предсказываемые значения функции отклика в каждой точке спектра плана \bar{y}_i . В уравнение регрессии при вычислениях подставляют значения нормированных факторов x_{ij} в соответствии с матрицей спектра плана. Затем определяется *остаточная дисперсия* $S_{\text{ост}}^2$, оценивающая погрешность полученной модели:

$$S_{\text{ост}}^2 = \frac{1}{N - N_B} \sum_{i=1}^N (y_i - \bar{y}_i)^2, \quad (11.93)$$

где N_B — число коэффициентов в уравнении регрессии.

Качество предсказания, обеспечиваемого полученной регрессионной моделью, оценивают по критерию Фишера F и коэффициенту детерминации R^2 .

При определении критерия Фишера принимается иная нулевая гипотеза, чем при экспериментах на вероятностных моделях (см. раздел 11.13). Здесь нулевая гипотеза гласит о том, что модель среднего $\bar{y}(\bar{X})$ достаточно хорошо описывает исследуемый процесс. Регрессионная модель окажется адекватной, если выдвинутая гипотеза будет опровергнута.

По критерию Фишера сравнивают дисперсии оцениваемой и противопоставляемой моделей. Последняя должна быть более точной, чем оцениваемая модель. Поэтому в данном случае критерий Фишера равен отношению дисперсии модели среднего S_y^2 к остаточной дисперсии $S_{\text{ост}}^2$:

$$F = S_y^2 / S_{\text{ост}}^2. \quad (11.94)$$

Уравнение регрессии адекватно описывает результаты эксперимента, если полученное по формуле (11.94) значение F больше табличного значения критерия Фишера F_T , определяемого при принятом уровне значимости q и числах степеней свободы k_1 и k_2 , с которыми определены дисперсии S_y^2 и $S_{\text{ост}}^2$. Согласно выраже-

ниям (11.92) и (11.93) $k_1 = N - 1$; $k_2 = N - N_B$. Если условие $F > F_T$ выполняется, это означает, что уравнение регрессии описывает результаты эксперимента в F_T раз лучше модели среднего. Тогда нулевая гипотеза отвергается и регрессионная модель адекватна.

Критерий детерминации определяется по формуле

$$R^2 = 1 - \frac{(N - N_B)S_{\text{ост}}^2}{(N - 1)S_y^2}. \quad (11.95)$$

Значение R^2 определяет долю рассеяния экспериментальных значений функции отклика, учитываемую регрессионной зависимостью. Модель считается работоспособной, если $R^2 \geq 0,75$.

При получении квадратичных регрессионных моделей оценку значимости коэффициентов обычно не производят и модель не упрощают, а используют полный квадратный полином со всеми его составляющими (свободным членом, линейными эффектами, эффектами парных взаимодействий и квадратичными членами).

Упрощению подвергаются лишь линейные многофакторные регрессии, полученные на основе ПФЭ. Принимается нулевая гипотеза о том, что $b_j = 0$, $j = \overline{1, N_B}$, и осуществляется проверка, отличаются ли статистически значимо оценки коэффициентов b_j от нуля. Значимость b_j проверяют по критерию Стьюдента, используя формулу (11.73). При вычислении дисперсии $S_{b_j}^2$, оценивающей погрешности определения коэффициентов b_j , используется остаточная дисперсия $S_{\text{ост}}^2$ (а не дисперсия воспроизводимости эксперимента S_y^2 , как это было для вероятностной модели (см. раздел 11.12)):

$$S_{b_j}^2 = S_{\text{ост}}^2 / N. \quad (11.96)$$

При уменьшении числа коэффициентов регрессии N_B , как видно из формулы (11.93), остаточная дисперсия возрастает, что приводит к снижению критерия Фишера. Поэтому члены уравнения регрессии с незначимыми коэффициентами b_j можно исключать лишь в том случае, если проверка полученной упрощенной модели на адекватность по критерию Фишера дает положительный результат.

**ПОЛУЧЕНИЕ МАТЕМАТИЧЕСКОЙ МОДЕЛИ
НА ОСНОВЕ ПАССИВНОГО
ЭКСПЕРИМЕНТА**

При проведении экспериментов на реальных технических объектах независимое варьирование факторов в большинстве случаев оказывается невозможным, поэтому для получения их математических моделей обычно проводятся пассивные эксперименты. Объекты при этом находятся в нормальных условиях функционирования, а изменение их фазовых координат и выходных параметров обусловлено влиянием внешних возмущающих воздействий, носящих случайный характер. В этой связи фазовые координаты и выходные параметры представляют собой случайные процессы. На них могут влиять изменения внутренних параметров объектов (например, параметров системы управления). Отнесем к факторам пассивного эксперимента $x_i, i = \overline{1, n}$ внешние воздействия, наблюдаемые фазовые координаты и изменяемые внутренние параметры, а к функциям отклика $y_j, j = \overline{1, m}$ — выходные параметры объекта (показатели качества и эффективности).

Для получения информации о физических свойствах объекта, необходимой при построении математической модели, выбирают некоторый интервал дискретизации независимой переменной (времени t) и фиксируют в дискретные моменты времени значения факторов и функций отклика. Эти значения представляют собой случайные последовательности чисел, составляющие непрерывные множества. Необходимо, чтобы эти случайные числа для каждого фактора и каждой функции отклика в отдельности были некоррелированными. Это достигается соответствующим выбором интервала дискретизации времени (см. раздел 10.9). Эксперимент должен проводиться таким образом, чтобы исследуемые случайные процессы были стационарными и эргодическими. Допускается лишь нестационарность по математическому ожиданию, которую можно легко отфильтровать и использовать центрированные значения случайных процессов.

Изложенный подход к проведению пассивного эксперимента составляет основу *метода статистических испытаний (Монте-Карло)*.

Используя полученные выборки факторов и функций отклика, находят их статистические оценки (см. раздел 10.9) и осуществляют построение регрессионной модели технического объекта $y_j = \varphi(\bar{X}, \bar{b})$, $j = \overline{1, m}$. При этом необходимо проверить выполнение 1-го и 5-го постулатов регрессионного анализа. Со-

гласно 1-му постулату все выборки должны иметь нормальные распределения (проверяется по критерию Пирсона). Затем проверяется выполнение 5-го постулата о некоррелированности столбцов матрицы факторов. Для этого осуществляется *корреляционный анализ* результатов статистических испытаний.

В процессе корреляционного анализа определяют оценки коэффициентов парной корреляции $r_{y_j x_i}$ между выбранными для построения математической модели выходными параметрами y_j и факторами x_i , а также между парами факторов x_i и x_k , т.е. оценки коррелированности этих факторов $r_{x_i x_k}$.

Оценка коэффициента корреляции между y_j и x_i вычисляется по формуле

$$r_{y_j x_i} = \frac{1}{(N-1)S_{y_j} S_{x_i}} \left(\sum_{u=1}^N y_{ju} x_{iu} - N \bar{y}_j \bar{x}_i \right), \quad i = \overline{1, n}; \quad j = \overline{1, m}, \quad (11.97)$$

где N — число проведенных опытов; y_{ju}, x_{iu} — значения переменных y_j и x_i в u -м опыте; \bar{y}_j, \bar{x}_i — оценки математических ожиданий (выборочные средние) соответственно функции отклика y_j и фактора x_i ; S_{y_j}, S_{x_i} — средние квадратические отклонения.

Коэффициенты $r_{y_j x_i}$ являются элементами корреляционной матрицы $R_{\bar{Y}\bar{X}} = (r_{ji})$, $i = \overline{1, n}$; $j = \overline{1, m}$, в которой $r_{ji} = r_{y_j x_i}$:

$$R_{\bar{Y}\bar{X}} = \begin{bmatrix} \bar{R}_1 \\ \bar{R}_2 \\ \dots \\ \bar{R}_j \\ \dots \\ \bar{R}_m \end{bmatrix} = \begin{bmatrix} r_{11} & r_{12} & \dots & r_{1n} \\ r_{21} & r_{22} & \dots & r_{2n} \\ \dots & \dots & \dots & \dots \\ r_{j1} & r_{j2} & \dots & r_{jn} \\ \dots & \dots & \dots & \dots \\ r_{m1} & r_{m2} & \dots & r_{mn} \end{bmatrix}, \quad (11.98)$$

где $\bar{R}_j = (r_{j1}, r_{j2}, \dots, r_{jn})$ — вектор-строка оценок коэффициентов парной корреляции между выходным параметром y_j и всеми факторами x_i , $i = \overline{1, n}$.

Аналогично вычисляются оценки коэффициентов парной корреляции $r_{x_i x_k}$ между факторами x_i и x_k :

$$r_{x_i x_k} = \frac{1}{(N-1)S_{x_i} S_{x_k}} \left(\sum_{u=1}^N x_{iu} x_{ku} - N \bar{x}_i \bar{x}_k \right), \quad i, k = \overline{1, n}. \quad (11.99)$$

Обозначим $r_{x_i x_k} = \rho_{ik}$. Тогда связь между факторами можно представить корреляционной матрицей

$$R_{\bar{X}} = \begin{bmatrix} \rho_{11} & \rho_{12} & \cdots & \rho_{1n} \\ \rho_{21} & \rho_{22} & \cdots & \rho_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ \rho_{n1} & \rho_{n2} & \cdots & \rho_{nn} \end{bmatrix}. \quad (11.100)$$

В матрице (11.100) $\rho_{ii} = 1$, $i = \overline{1, n}$, а остальные коэффициенты корреляции могут принимать значения в пределах $0 \leq |\rho_{ik}| \leq 1$. Если $|\rho_{ik}|$ близко к 1, это свидетельствует о сильной коррелированности факторов x_i и x_k , а при $|\rho_{ik}| = 1$ эти факторы функционально (не вероятно) связаны между собой. Оценка влияния каждого из них на функцию отклика по уравнению регрессии окажется невозможной. В случае сильной корреляции факторов x_i и x_k один из них следует исключить. Для построения уравнения регрессии $y_j = \varphi(\bar{X}, \vec{b})$ оставляют тот фактор, у которого коэффициент корреляции r_{u, x_i} больше.

Регрессионный анализ результатов пассивного эксперимента выполняется по той же методике, что и активного. Факторы нормируют с использованием формул (11.1) — (11.3). Но в пассивном эксперименте значения факторов — случайные числа, поэтому после нормирования каждый из них во всей серии опытов распределяется в диапазоне $-1 \leq x_i \leq +1$. В результате матрица Фишера не диагональная, как в активном эксперименте, а может иметь все ненулевые элементы. Поэтому определение коэффициентов регрессии b_j требует формирования и решения системы алгебраических уравнений (11.20).

Оценка качества предсказания, обеспечиваемого полученной моделью, осуществляется по критерию Фишера F и по коэффициенту детерминации R^2 , вычисляемых соответственно по формулам (11.94) и (11.95).

Если необходимо получить линейное уравнение регрессии в виде (11.26), можно использовать корреляционные матрицы (11.98) и (11.100).

Используя вектор-строку \bar{R}_j матрицы (11.98) для j -го выходного параметра y_j и корреляционную матрицу (11.100), составим матричное уравнение

$$R_{\bar{X}} \bar{H}_j = \bar{R}_j, j = \overline{1, m}. \quad (11.101)$$

Решив систему линейных алгебраических уравнений (11.101) относительно неизвестного вектора $\bar{H}_j = (h_{ji}), i = \overline{1, n}$, вычислим искомые коэффициенты уравнения регрессии (11.26) по формуле

$$b_{ji} = h_{ji} S_{y_j} / S_{x_i}. \quad (11.102)$$

Если факторы x_i — независимые случайные величины, то матрица $R_{\bar{X}}$ единичная, т.е. $\rho_{ii} = 1, \rho_{ik} = 0, i \neq k$. Тогда $h_{ji} = r_{ji}$.

Изложенную методику можно также использовать при решении задач прогнозирования параметров технических объектов. На основе ретроспективного анализа составляют матрицы важнейших внутренних и выходных параметров объектов данного типа. Такой анализ аналогичен проведению пассивного эксперимента. Сущность эксперимента в данном случае состоит в создании множества конструкций одного и того же целевого назначения. Собранные информация об объекте представляет собой статистические выборки – конечные наборы значений реализаций случайных величин (факторов и функций отклика). Регрессионная модель, полученная в результате такого эксперимента, позволяет осуществлять поиск оптимальных решений и прогнозирование. При этом допускается экстраполяция параметров с выходом за пределы факторного пространства.