Лекция 6

Планирование полного факторного эксперимента

В методе полного факторного эксперимента (ПФЭ) ограничиваются линейной частью разложения и членами, содержащими произведения факторов в первой степени (учитываются линейные и гиперболические слагаемые). Этот способ оказывается наиболее предпочтительным в тех случаях, когда отсутствует априорная информация для обоснования структуры модели с позиций физических представлений процессов, происходящих в объекте, отсутствует количественная оценка степени влияния изучаемых факторов на выходную переменную объекта, его выходной показатель.

Коэффициенты искомого уравнения определяют на основе экспериментальных данных и, следовательно, несут на себе отпечаток погрешностей эксперимента. Чтобы подчеркнуть это обстоятельство, в уравнении вместо символов β, обозначающих истинные значения коэффициентов, пишут b, подразумевая под этим соответствующие выборочные оценки.

Итак, с помощью ПФЭ ищут математическое описание процесса в виде уравнения

$$y = b_0 + b_1x_1 + b_2x_2 + b_3x_3 + b_4x_2x_1 + b_5x_2x_3 + b_6x_3x_1 \quad (1)$$

Его называют уравнением регрессии, а коэффициенты – коэффициентами регрессии.

Для удобства вычислений коэффициентов регрессии все факторы в ходе полного факторного эксперимента варьируют на двух уровнях, соответствующих значениям кодированных переменных +1 и –1. Таким образом, *полным факторным экспериментом* называется система опытов, содержащая все возможные неповторяющиеся комбинации верхнего и нижнего уровней факторов.

Приведем этапы планирование и реализации полного факторного эксперимента.

1-этап. Выбираем параметры оптимизации (аргументы функции цели) и уровень их варьирования (диапазон изменения).

2-этап. Далее происходит переход к другой, более удобной системе координат. Этот процесс называется кодированием.

3-этап. Составления матрицы планирования эксперимента.

4-этап. Рандомизация. Рандомизауция опыта. Она нужна для того что бы избавится от предвзятых измерений. На пример в каких то измерения появилась ошибка (детерминированная, систематическая) – то есть какое-то предвзятое отношение к какой-то точке измерения. После рандомизации происходит размазывание этой ошибки.

5-этап. Реализация плана эксперимента.

6-этап. Проверка однородности дисперсии параллельных опытов, воспроизводимости результатов.

7-этап. Расчет коэффициентов уравнения регрессии, их ошибок и значимости.

8-этап. Проверка адекватности модели.

Приведем ***пример кодирования.***

Факторы	x_1	X_1	x_2	X_2
Интервал варьирования	0,75	1	1	1
Верхний уровень	2,5	+1	3	+1
Нижний уровень	1	-1	1	-1
Основной уровень (середина)	1,75	0	2	0

$x_1 := (0.75\ \ 2.5\ \ 1\ \ 1.75)^T$

Произведем действия над первой величиной x_1 (первый фактор)

$x_{1min} := 1 \qquad x_{1max} := 2.5$

$\Delta x := \dfrac{x_{1max} - x_{1min}}{2} = 0.75$ Интервал варьирования фактора

$x_{10} := \dfrac{x_{1max} + x_{1min}}{2} = 1.75$ Нулевой уровень (середина) - Основной уровень (фактора)

$X1 := \dfrac{x_{1max} - x_{10}}{\Delta x} = 1$ Кодирование значения верхнего уровня

$X2 := \dfrac{x_{1min} - x_{10}}{\Delta x} = -1$ Кодирование значения нижнего уровня

$X_{10} := \dfrac{x_{10} - x_{10}}{\Delta x} = 0$ Кодированный нулевой уровень - Основной уровень (фактора)

$x_2 := (1\ \ 3\ \ 1\ \ 2)^T$

Произведем действия над второй величиной x_2 (второй фактор). Они аналогичны. Приведем программу

$$\mathrm{Kod}(x) := \left| \begin{array}{l} x_{max} \leftarrow \max(x) \\ x_{min} \leftarrow \min(x) \\ \Delta \leftarrow \dfrac{x_{max} - x_{min}}{2} \\ x_0 \leftarrow \dfrac{x_{max} + x_{min}}{2} \\ x1 \leftarrow \dfrac{x_{max} - x_0}{\Delta} \\ x2 \leftarrow \dfrac{x_{min} - x_0}{\Delta} \\ xo \leftarrow 0 \\ \begin{pmatrix} \Delta & \dfrac{\Delta}{\Delta} \\ x_{max} & x1 \\ x_{min} & x2 \\ x_0 & xo \end{pmatrix} \end{array} \right.$$

$x_1 := (2.5\ \ 1)^T \qquad x_2 := (3\ \ 1)$

$$X1 := \mathrm{Kod}(x_1) = \begin{pmatrix} 0.75 & 1 \\ 2.5 & 1 \\ 1 & -1 \\ 1.75 & 0 \end{pmatrix} \qquad X2 := \mathrm{Kod}(x_2) = \begin{pmatrix} 1 & 1 \\ 3 & 1 \\ 1 & -1 \\ 2 & 0 \end{pmatrix}$$

Ниже приводятся графические зависимости факторов в разных системах координат

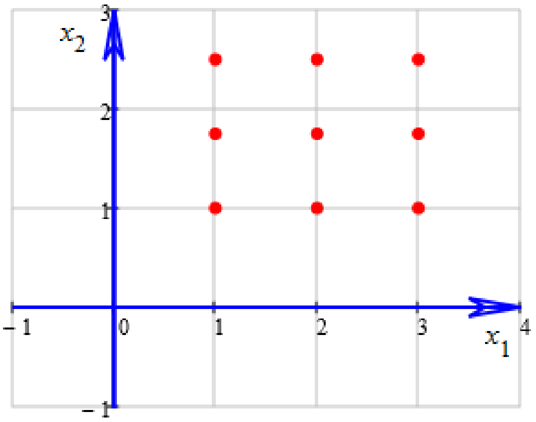

Рис. 1. Факторное пространство. Исходные значения эксперимента

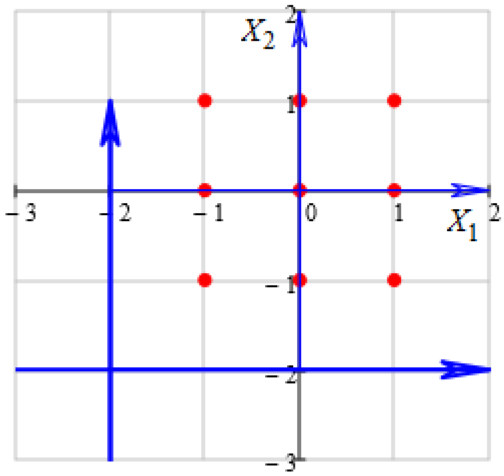

Рис. 2. Факторное пространство. Кодированные значения эксперимента (Новая система координат)

Если число факторов три, то они кодируются в виде таблички приведенной ниже. Для первого фактора x_1 единицы чередуются знаками. Для второго фактора x_2 единицы идут в таком порядке: две единицы положительные, две отрицательные и т.д. Для третьего фактора x_3 первая половина единиц отрицательна, вторая положительна. Как в двоичных кодах.

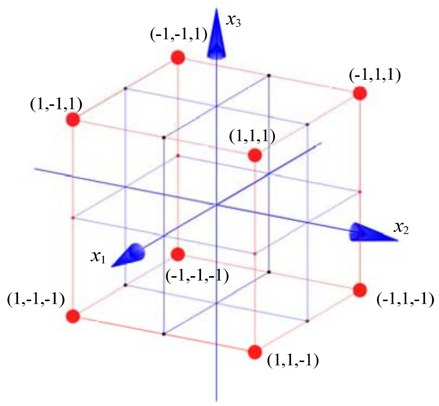

Рис. 3. Факторное пространство. Кодирование для трёхфакторного эксперимента

T =

	0	1	2
0	"x1"	"x2"	"x3"
1	-1	-1	-1
2	1	-1	-1
3	-1	1	-1
4	1	1	-1
5	-1	-1	1
6	1	-1	1
7	-1	1	1
8	1	1	1

Рис. 4. Таблица кодирование для трёхфакторного эксперимента

Матрица планирования эксперимента

Условия эксперимента записывают в виде матрицы планирования эксперимента. Где строки соответствуют различным независимым испытаниям.

Тб. Матрица планирования эксперимента 2^2

Номер испытания	X_1	X_2	y
1	-1	-1	y_1
2	1	-1	y_2
3	-1	1	y_3
4	1	1	y_4

В общем случае планы типа 2^k, в общем случае представляет собой совокупность точек расположенных в вершинах гиперкуба в многомерном пространстве. Пространство, которое находится внутри куба, является областью планирования эксперимента (множество точек внутри куба). Один из возможных методов формирования таблицы – это чередование знаков, которое было описано выше. Графики для 2^2 , 2^3 приведены на рис. 2, рис.3.

Номер опыта	X_1	X_2	X_3	X_4
1	-	-	-	-
2	+	-	-	-
3	-	+	-	-
4	+	+	-	-
5	-	-	+	-
6	+	-	+	-
7	-	+	+	-
8	+	+	+	-
9	-	-	-	+
10	+	-	-	+
11	-	+	-	+
12	+	+	-	+
13	-	-	+	+
14	+	-	+	+
15	-	+	+	+
16	+	+	+	+

Такой тип построения планов является наиболее эффективным для построения линейных моделей. Эффективность достигается за счет следующих свойств:

1. *Симметрия относительно центра эксперимента.* Алгебраическая сумма всех значений в столбцах равна нулю (его можно назвать условием ортогональности)

$$\sum_{k=0}^{length(x1)-1} x1_k = 0 \quad \sum_{k=0}^{length(x2)-1} x2_k = 0 \quad \sum_{k=0}^{length(x3)-1} x3_k = 0$$ для Рис. 4

2. *Условие нормировки,* когда сумма квадратов элементов таблицы равна числу опытов. Это следствие того что все значения в таблице равны +1 или -1.

$$\sum_{k=0}^{\text{length}(x1)-1} \left(x1_k\right)^2 = 8 \qquad \sum_{k=0}^{\text{length}(x2)-1} \left(x2_k\right)^2 = 8 \qquad \sum_{k=0}^{\text{length}(x3)-1} \left(x3_k\right)^2 = 8$$

3. *Ортогональность.* О ней уже было упомянуто в пункте 1. Сумма произведений элементов двух любых столбцов равна нулю. Это свойство позволяет легко определять регрессионные коэффициенты разложения целевой функции. То есть для каждого коэффициента разложения получается независимая формула.

$$\sum_{k=0}^{\text{length}(x1)-1} \left(x1_k \cdot x3_k\right) = 0 \qquad \sum_{k=0}^{\text{length}(x1)-1} \left(x2_k \cdot x3_k\right) = 0 \qquad \sum_{k=0}^{\text{length}(x1)-1} \left(x1_k \cdot x2_k\right) = 0$$

Например, для случая

$$y = b_0 + b_1x_1 + b_2x_2 + b_3x_3 + b_4x_2x_1 + b_5x_2x_3 + b_6x_3x_1$$

Имеем

Номер опыта	X_1	X_2	X_3	y
1	-	-	-	y_1
2	+	-	-	y_2
3	-	+	-	y_3
4	+	+	-	y_4
5	-	-	+	y_5
6	+	-	+	y_6
7	-	+	+	y_7
8	+	+	+	y_8

$$\sum_{k=1}^{N} y_k = b_0N \;\rightarrow\; b_0 = \frac{1}{N}\sum_{k=1}^{N} y_k$$

$$\sum_{k=1}^{N} y_kx_{1k} = b_1N \;\rightarrow\; b_1 = \frac{1}{N}\sum_{k=1}^{N} y_kx_{1k}\,, \qquad \sum_{k=1}^{N} y_kx_{2k} = b_2N \;\rightarrow\; b_2 = \frac{1}{N}\sum_{k=1}^{N} y_kx_{2k}\,,$$

$$\sum_{k=1}^{N} y_kx_{3k} = b_3N \;\rightarrow\; b_3 = \frac{1}{N}\sum_{k=1}^{N} y_kx_{3k}$$

$$\sum_{k=1}^{N} y_kx_{1k}x_{2k} = b_4N \;\rightarrow\; b_4 = \frac{1}{N}\sum_{k=1}^{N} y_kx_{1k}x_{2k}\,, \qquad \sum_{k=1}^{N} y_kx_{2k}x_{3k} = b_5N \;\rightarrow\; b_5 = \frac{1}{N}\sum_{k=1}^{N} y_kx_{2k}x_{3k}$$

$$\sum_{k=1}^{N} y_kx_{1k}x_{3k} = b_6N \;\rightarrow\; b_6 = \frac{1}{N}\sum_{k=1}^{N} y_kx_{1k}x_{3k}$$

Благодаря свойству ортогональности процедура определения коэффициентов превращается в простую арифметическую процедуру.

В частности для двух факторного эксперимента получаем

$$y = b_0 + b_1x_1 + b_2x_2 + b_3x_2x_1$$

Номер испытания	X_1	X_2	y
1	-1	-1	y_1
2	1	-1	y_2
3	-1	1	y_3
4	1	1	y_4

$$\sum_{k=1}^{N} y_k = b_0 N \;\rightarrow\; b_0 = \frac{1}{N}\sum_{k=1}^{N} y_k = \frac{1}{4}\left(y_1 + y_2 + y_3 + y_4\right)$$

$$\sum_{k=1}^{N} y_k x_{1k} = b_1 N \;\rightarrow\; b_1 = \frac{1}{N}\sum_{k=1}^{N} y_k x_{1k} = \frac{1}{4}\left(y_1 - y_2 + y_3 - y_4\right)$$

$$\sum_{k=1}^{N} y_k x_{2k} = b_2 N \;\rightarrow\; b_2 = \frac{1}{N}\sum_{k=1}^{N} y_k x_{2k} = \frac{1}{4}\left(-y_1 - y_2 + y_3 + y_4\right)$$

$$\sum_{k=1}^{N} y_k x_{1k} x_{2k} = b_3 N \;\rightarrow$$

$$b_3 = \frac{1}{N}\sum_{k=1}^{N} y_k x_{1k} x_{2k} = \frac{1}{4}\left(y_1(-1)\cdot(-1) + y_2(1)\cdot(-1) - y_3(-1)\cdot(1) + y_4(1)\cdot(1)\right) =$$

$$= \frac{1}{4}\left(y_1 - y_2 - y_3 + y_4\right)$$

4. *Ротатабельность.* Точки расположены так, что точность предсказания параметра оптимизации одинаково на равных расстояниях от центра плана и не зависит от направления.

Рандомизация эксперимента

Что бы исключить влияние систематических погрешностей (субъективность, предвзятость), вызванных внешними условиями, применяется метод рандомизации (случайность). Идея состоит в переводе систематических погрешностей в случайные погрешности. Например, если в опыте проведено 2^3 опытов, то нужно 16 опытов. При этом пользуются таблицей случайных чисел (или генератором случайных чисел). Выбранную последовательность чисел не следует нарушать.