УДК 519.681

ОЦЕНКА ПРЕДСТАВИТЕЛЬНОСТИ УСЕЧЕННЫХ ОРТОГОНАЛЬНЫХ ПОДПЛАНОВ ПЛАНА ПОЛНОГО ФАКТОРНОГО ЭКСПЕРИМЕНТА

О.В. СЕРАЯ, Д.А. ДЕМИН

Введен критерий оценки качества усеченного ортогонального подплана плана полного факторного эксперимента, характеризующий уровень смешивания влияний факторов и их взаимодействий. Предложена методика расчета степени представительности усеченных ортогональных планов. Учет введенного критерия позволяет повысить точность оценивания параметров многооператорной регрессии в условиях малой выборки.

ВВЕДЕНИЕ

При решении большого числа разнообразных задач в технике, экономике, социологии и т.д. традиционно возникает проблема анализа системы типа «состав – свойство». Общепринятый и широко используемый подход к решению этой проблемы состоит в выявлении и отборе совокупности факторов, предположительно влияющих на значение некоторой функции отклика, характеризующей эффективность функционирования системы. Эта функция отклика обычно формализуется в виде нелинейного по факторам, но линейного по параметрам уравнения регрессии. Задача состоит в оценивании параметров этого уравнения путем статистической обработки результатов пассивного эксперимента. Характерная трудность решения практических задач этого типа в условиях большого числа влияющих факторов состоит в недостаточности экспериментальных данных. Описанный в [1] подход позволяет осуществить искусственную ортогонализацию пассивного эксперимента, что потенциально обеспечивает возможность независимого оценивания степени влияния факторов и их взаимодействий. При этом в случае неоднородности плана полного факторного эксперимента (ПФЭ) в смысле ошибок измерения значений функции откликов [1] предложена методика построения усеченного ортогонального подплана ПФЭ.

Построение усеченного представительного ортогонального плана производится следующим образом. Пусть количество факторов равно m. Разобьем эту совокупность факторов на три подмножества $\{A,B,C\}$ по $p=\dfrac{m}{3}$ факторов в каждом, например,

$$A=\{F_1,F_2,...,F_p\},\quad B=\{F_{p+1},F_{p+2},...,F_{2p}\},\quad C=\{F_{2p+1},F_{2p+2},...,F_{3p}\}.$$

Далее введем независимую нумерацию факторов в этих подмножествах.

Пусть

$$F_l^A = F_l, \; F_l^B = F_{p+l}, \; F_l^C = F_{2p+l}, \; l = 1, 2, ..., p .$$

Сформируем ортогональный полный m-факторный план. Общее число экспериментов плана равно 2^m, а количество экспериментов, соответствующих всем возможным комбинациям уровней для факторов, входящих в любое из подмножеств, равно 2^p. Пронумеруем эти комбинации для факторов F_l^A, $l = 1,2,...,p$ индексом i_1, а для факторов подмножеств F_l^B, $l = 1,2,...,p$ и F_l^C, $l = 1,2,...,p$ соответственно индексами i_1 и i_2. Тогда любой строке плана может быть поставлена во взаимнооднозначное соответствие тройка ($i_1\, i_2\, i_3$).

Множество комбинаций индексов $\{i_1\, i_2\, i_3\}$ может быть представлено в виде трехмерной решетки с 2^p узлами по каждой из трех координат. Введем параметр

$$Z_{i_1 i_2 i_3} = \begin{cases} 1, \text{ если строка } (i_1, i_2, i_3) \text{ включена в эксперимент}, \\ 0 \quad \text{в противном случае}. \end{cases}$$

Понятно, что набор значений $\{Z_{i_1 i_2 i_3}\}$, $i_1 = 1,2,...,2^p$, $i_2 = 1,2,..,2^p$, $i_3 = 1,2,...,2^p$, однозначно определяет некоторый план эксперимента, содержащий число строк, равное числу единиц в наборе. Введем теперь следующую систему уравнений с булевыми переменными

$$\sum_{i_3=1}^{2^p} Z_{i_1 i_2 i_3} = 1, \; i_1 = 1,2,...,2^p, \; i_2 = 1,2,..,2^p, \tag{1}$$

$$\sum_{i_2=1}^{2^p} Z_{i_1 i_2 i_3} = 1, \; i_1 = 1,2,...,2^p, \; i_2 = 1,2,..,2^p, \tag{2}$$

$$\sum_{i_1=1}^{2^p} Z_{i_1 i_2 i_3} = 1, \; i_1 = 1,2,...,2^p, \; i_2 = 1,2,..,2^p. \tag{3}$$

Как доказано в [1], любое решение системы уравнений (1)–(3) определяет симметричный ортогональный план эксперимента с числом строк, равным 2^{2p}.

ПОСТАНОВКА ЗАДАЧИ

Выбранное решение системы уравнений (1)–(3) однозначно задает выделяемую совокупность строк полного факторного эксперимента, которые образуют ортогональный усеченный подплан ПФЭ. Отметим, что различные подпланы, соответствующие разным решениям системы уравнений (1)–(3), отличаются уровнем представительности, который вычисляется по формуле:

$$F_1 = \sum_{i_1=1}^{2^p} \sum_{i_2=1}^{2^p} \sum_{i_3=1}^{2^p} \sigma^2_{i_1 i_2 i_3} Z_{i_1 i_2 i_3} \,, \tag{4}$$

или по формуле

$$F_2 = \max_{i_1 i_2 i_3} \{\sigma^2_{i_1 i_2 i_3} Z_{i_1 i_2 i_3}\} \,, \tag{5}$$

где $\sigma^2_{i_1 i_2 i_3}$ — дисперсия оценки значения функции отклика в вершине $(i_1\, i_2\, i_3)$.

При этом соотношение (4) определяет суммарную дисперсию ошибок оценивания значений функции отклика в ортогональных точках плана, а в соотношении (5) — максимальную дисперсию. Вместе с тем, важно учитывать, что есть еще один существенный критерий, сегрегирующий разные планы, смысл которого состоит в следующем. Как показано в [1], решение системы (1)–(3) обеспечивает ортогональность факторов, входящих в уравнение регрессии (1). Однако, взаимодействия факторов в выделенном подплане необязательно ортогональны факторам и между собой. Понятно, что если столбец значений какого-либо взаимодействия не ортогонален столбцу значений какого-либо фактора или какого-либо другого взаимодействия, то скалярное произведение этих столбцов примет значение из следующего набора: $\{2^p, 2^{p+1}, \ldots, 2^{2p}\}$.

При этом независимое оценивание влияния на функцию отклика каждого из компонентов этого скалярного произведения невозможно, эти влияния смешиваются. Отсюда следует, что степень предпочтительности подплана будет тем более высокой, чем меньшее число взаимодействий факторов не ортогональны факторам и между собой.

РЕЗУЛЬТАТЫ

Не снижая общности, технологию формирования представительного усеченного ортогонального подплана опишем на конкретном примере. Пусть в результате искусственной ортогонализации сформирован ПФЭ для шести факторов, содержащий $2^6 = 64$ эксперимента.

В соответствии с методикой разобьем совокупность факторов на три группы по $p = 2$ фактора в каждой. При этом количество экспериментов, соответствующих всем возможным комбинациям уровней для факторов, входящих в любое подмножество, равно $2^p = 2^2 = 4$. Перенумеруем эти комбинации для факторов первого подмножества индексом i_1, второго — индексом i_2 и третьего — индексом i_3. Теперь любой строке ПФЭ можно поставить в соответствие тройки $(i_1\, i_2\, i_3)$, $i_1 \in \{1, 2, 3, 4\}$, $i_2 \in \{1, 2, 3, 4\}$, $i_3 \in \{1, 2, 3, 4\}$

Трехмерную решетку $(i_1\, i_2\, i_3)$ удобно представить в виде четырех планарных сечений следующим образом (рисунок).

На рисунке набором единиц задано одно из возможных решений системы уравнений (1)–(3). Легко найти множество строк, соответствующих какому-либо решению системы (1)–(3) и образующих подплан ПФЭ. Для этого необходимо каждой компоненте набора $\{Z_{i_1 i_2 i_3}\}$, являющегося решением системы уравнений (1)–(3), поставить в соответствие номер строки, вычисляемой по формуле

$$l(i_1 i_2 i_3) = (i_1 - 1)2^{2p} + (i_2 - 1)2^p + (i_3 - 1) .$$

При этом, например компоненте $Z_{111} = 1$ соответствует номер строки $l(111) = 0$, а компоненте $Z_{443} = 1$ — номер строки $l(443) = 48 + 12 + 2 = 62$. Перечислим все множество номеров строк, выделяемых решением, представленным на рисунке.

$$L = \{1; 6; 11; 16; 18; 23; 28; 29; 35; 40; 41; 46; 52; 53; 58; 63\} .$$

$i_1 = 1$

i_2 \ i_3	1	2	3	4
1	**1**			
2		**1**		
3			**1**	
4				**1**

1

$i_1 = 2$

i_2 \ i_3	1	2	3	4
1		**1**		
2			**1**	
3				**1**
4	**1**			

4

$i_1 = 3$

i_2 \ i_3	1	2	3	4
1			**1**	
2				**1**
3	**1**			
4		**1**		

2 3

$i_1 = 4$

i_2 \ i_3	1	2	3	4
1				**1**
2	**1**			
3		**1**		
4			**1**	

Рисунок. Одно из решений системы (1)–(3)

Теперь для этого подплана рассчитываем скалярные произведения векторов-столбцов, соответствующих факторам, и их взаимодействий. При этом, в частности, оказывается, что

$$\langle F_1, F_{35} \rangle = -8,\ \langle F_2, F_{46} \rangle = -16,\ \langle F_1, F_{235} \rangle = 8,\ \langle F_1, F_{345} \rangle = 8 .$$

Это означает, что влияние фактора F_1 смешивается с влиянием взаимодействия F_{35}, влияние фактора F_2 смешивается с влиянием взаимодействия F_{46} и т.д. Совокупность пар неортогональных векторов-столбцов определяют качество выбранного подплана. Важно заметить, что выявляемые в результате описанной процедуры свойства подплана никак не связаны с результатами эксперимента и поэтому для каждого возможного решения системы могут быть проанализированы заранее.

Общее число возможных решений системы уравнений (1)−(3), очевидно, равно $N=(2^p)!(2^{p-1})!\ldots 2!$. Если $p=2$, то $N=4!3!2!=288$. Степень предпочтительности плана тем выше, чем выше порядок взаимодействия факторов, смешиваемых с основными факторами, и чем меньше общее число пар неортогональных столбцов.

ВЫВОДЫ

Таким образом, традиционный набор критериев, используемый обычно при выборе одного из множества возможных усеченных ортогональных подпланов плана полного факторного эксперимента, дополнен еще одним, существенно характеризующим качество усеченного подплана. Введенный критерий определяет уровень смешивания влияний факторов и их взаимодействий на значение функции отклика.

ЛИТЕРАТУРА

1. *Серая О.В., Демин Д.А.* Оценивание параметров уравнения регрессии в условиях малой выборки // Восточно-Европейский журнал передовых технологий. — 2009. — № 6/4(42). — С. 14–19.

Поступила 02.02.2010