

ПОСТРОЕНИЕ МОДЕЛИ ОБЪЕКТА УПРАВЛЕНИЯ ПО РЕЗУЛЬТАТАМ ПАССИВНОГО ЭКСПЕРИМЕНТА

Федорченко С.Г.¹

(Приднестровский государственный университет)

В статье рассмотрено развитие модифицированного метода случайного баланса, позволяющего построить математические модели высокого порядка по результатам пассивного эксперимента.

Пассивный эксперимент, нормальный закон распределения, ортогональность факторов, опорные области

1. Модифицированный метод случайного баланса для пассивного эксперимента

Для обработки результатов пассивного эксперимента можно использовать модифицированный метод случайного баланса для пассивного эксперимента (ММСБП) [1, 4].

Результатом пассивного эксперимента является таблица, каждая строка которой представляет собой числовое значение целевой функции (выходного показателя качества исследуемого объекта) при некоторых условиях (в определенный момент времени или для определенной партии изделий, или при прохождении определенной технологической операции и т.п.) и числовые значения исследуемых факторов при тех же условиях. Такая таблица, как правило, строится в результате длительных контрольных измерений выходного показателя качества однородной продукции и сопутствующих ему факторов, например режимов технологических операций или параметров самого изделия на предстоящих операциях.

¹ Федорченко Сергей Григорьевич, кандидат технических наук, доцент (fed_tir@mail.ru)

Пусть на исследуемый процесс (объект) воздействуют факторы $x_1, x_2, x_3, \dots, x_k$, тогда искомая модель примет вид $\hat{y} = b_0 + b_1x_1 + b_2x_2 + \dots + b_kx_k$.

Как известно, общими требованиями всех факторных планов являются нормальность закона распределения целевой функции Y , некоррелированность (слабая коррелированность), ортогональность и гомоскедастичность факторов X_k (равенство выборочных дисперсий во всех точках факторного пространства).

Ортогональность факторов в активном эксперименте обеспечивается переносом центра координат в базовую точку эксперимента и выбором точек факторного пространства, расположенных в вершинах гиперкуба (длина каждого ребра принимается за удвоенную новую единицу измерения), вписанного в гиперсферу определенного радиуса.

В пассивном эксперименте имеет место лишь естественное производственное варьирование значений факторов, как правило, в пределах допуска на фактор, т.е. незначительное. Следовательно изменение целевой функции также будет небольшим, и чтобы отличить его от шумовых флуктуаций, необходимо иметь достаточно длинную таблицу, в которой возможный эффект воздействия конкретного параметра на целевую функцию проявился бы в достаточной мере. Экспериментальным путем установлено, что таблица результатов пассивного эксперимента будет достаточно длинной, если на каждый исследуемый в ней фактор приходится 10-15 строк, но не более 350 строк всего.

Для нахождения модели с помощью ММСБП данные из таблицы исходных данных необходимо перевести в новую систему координат в относительных единицах. С этой целью и для увеличения точности результатов будущих расчетов весь диапазон $X_{k \max} - X_{k \min}$ каждого фактора X_k следует разбить на три части таким образом, чтобы число попаданий в каждую из них было примерно одинаковым, при этом части следует кодировать символами -1, 0 и +1. Хотя вид закона распределения факторов не оговаривается, для них можно оговорить правило: все значения $X_k \leq \bar{X}_k - 0,5S_k$ будут относиться к области $x_k = -1$, все значения $X_k \leq \bar{X}_k + 0,5S_k$ - к области $x_k = +1$, а остальные зна-

чения X_k - к области $x_k = 0$ (здесь \bar{X}_k - среднее арифметическое, S_k - среднеквадратическое отклонение числового значения фактора X_k , определенное выборке, причем символом X_k обозначаются значения k -го фактора в абсолютных единицах, а x_k - в относительных).

В результате исходная таблица превращается в план квази-активного эксперимента. При переходе к относительным координатам толщина оболочки гиперсферы увеличивается (или, иначе, вместо точечных вершин гиперкуба появляются некоторые «вершинные» области). Поэтому требуемая точность выделения значимых факторов и определения оценок коэффициентов регрессии может быть обеспечена только за счет увеличения числа опытов.

Исходную таблицу данных следует превратить в план эксперимента, заменяя числовые значения X_{kj} относительными x_{kj} с помощью преобразования координат. Поскольку таблица исходных данных достаточно длинная, неизбежны совпадения некоторых строк плана, у каждой из которых имеется свое значение выходной величины. Такие совпадающие строки плана следует совместить, т.е. представить в конечном плане в виде одной строки с несколькими значениями выходной величины, которые необходимо рассматривать как выборку. Из плана эксперимента можно извлечь информацию о влиянии парных взаимодействий. С этой целью в план эксперимента включаются столбцы парных взаимодействий, каждая координата которых получается простым перемножением кодов координат исходных факторов.

Экспериментальные данные, как правило, содержат некоторое количество «грубых промахов», не присущих исследуемому объекту. Большинство этих грубых промахов не видны на общем фоне, однако с расслоением общей выборки на частные по строкам плана, появляется возможность проверить каждую строчную выборку на однородность (отсутствие «грубых промахов») любым из известных способов.

Для оставшихся после устранения «грубых промахов» данных плана вычисляются средние арифметические и дисперсии по каждой строке. Для проверки воспроизводимости опытных данных, применяется критерий Бартлетта [2, с.29]. При этом

средняя дисперсия опытов может быть принята равной средневзвешенной $S^2\{Y\} = S_p^2$.

Для расчетов оценок коэффициентов регрессии b_k и их дисперсии D_k следует использовать специальные выражения, учитывающие поправки на это нарушение гомоскедастичности и являющиеся в этих условиях более эффективными, чем другие оценки:

$$(1) \quad b_k = \frac{\left(\frac{D_{2k}}{N_{2k}} + 2m_k^2\right) \cdot \mu_{1k} - \left(\frac{D_{1k}}{N_{1k}} + 2m_k^2\right) \cdot \mu_{2k}}{\frac{D_{1k}}{N_{1k}} + \frac{D_{2k}}{N_{2k}} + 4m_k^2};$$

$$(2) \quad D_k = \frac{\frac{D_{1k}}{N_{1k}} \cdot \frac{D_{2k}}{N_{2k}} + \left(\frac{D_{1k}}{N_{1k}} + \frac{D_{2k}}{N_{2k}}\right) \cdot m_k^2}{\frac{D_{1k}}{N_{1k}} + \frac{D_{2k}}{N_{2k}} + 4m_k^2};$$

где $\mu_{1k} = \frac{1}{N_{1k}} \sum_{j=1}^{N_{1k}} Y_j^{(1k)}$; $\mu_{2k} = \frac{1}{N_{2k}} \sum_{j=1}^{N_{2k}} Y_j^{(2k)}$;

$\{Y_j^{(1k)}\}_{N_{1k}}$ и $\{Y_j^{(2k)}\}_{N_{2k}}$ - подмножества элементов выходной

величины из общей выборки, для которых x_{kj} имеет соответственно положительный и отрицательный знак;

N_{1k}, N_{2k} - объем соответствующих подмножеств;

$N_k = N_{1k} + N_{2k}$ - общий объем выборки для k -го фактора;

$m_k = \frac{1}{N_k} (\mu_{1k} N_{1k} + \mu_{2k} N_{2k})$ - оценка математического ожидания;

$$D_{1k} = \frac{1}{N_{1k} - 1} \sum_{j=1}^{N_{1k}} (Y_j^{(1k)} - \mu_{1k})^2; D_{2k} = \frac{1}{N_{2k} - 1} \sum_{j=1}^{N_{2k}} (Y_j^{(2k)} - \mu_{2k})^2$$

- дисперсии выходной величины соответственно при положительных и отрицательных значениях фактора x_k .

Значимость каждой полученной оценки коэффициента регрессии можно определить с помощью выражения (3). При выполнении условия

$$(3) \quad t_k = \frac{|b_k|}{\sqrt{D_k}} \geq t_{\text{табл}}(q; \nu_k)$$

где q – уровень значимости;

$\nu_k = N_k - 2$ - число степеней свободы;

$t_{\text{табл}}(q, \nu)$ – табличное значение распределения Стьюдента, оценки b_k признаются значимыми и должны быть включены в математическую модель.

План эксперимента и расслоение данных по строкам плана после преобразования выходной величины являются окончательными. Если среди незначимых оказался хотя бы один основной фактор, то вследствие его исключения план эксперимента нужно скорректировать и всю работу по определению оценок коэффициентов и их значимости проделать заново.

Для проверки найденной модели на адекватность необходимо найти дисперсию адекватности:

$$(4) \quad S_{\text{ад}}^2 = \frac{1}{\sum_{j=1}^N m_j - d} \sum_{j=1}^N m_j (\bar{Y}_j - \hat{Y}_j)^2,$$

где d – число членов найденной модели, N - количество строк плана.

При сравнении дисперсии неадекватности со средневзвешенной дисперсией опытов по критерию Фишера:

$$(5) \quad F = \frac{S_{\text{ад}}^2}{S_p^2} \leq F_{\text{табл}} \left(q; \nu_{\text{числ}} = \sum_{j=1}^N m_j - d; \nu_{\text{зн}} = \nu_z \right).$$

В случае удовлетворения неравенства (5) считается доказанным что полученная модель не противоречит опытным данным.

2. Построение моделей высоких порядков

2.1 Уравнение связи 2-го порядка

В ММСБП предложена кодировка факторов, графически представленная на рис. 1. Область определения каждого фактора разбивается на три области с границами X_n и X_e ($X_H = \bar{X} - 0,5S$

; $X_B = \bar{X} - 0,5S$), значения фактора в этих областях кодируются соответственно как -1, 0, +1 и уже относительно кодированных значений факторов строится математическая модель.

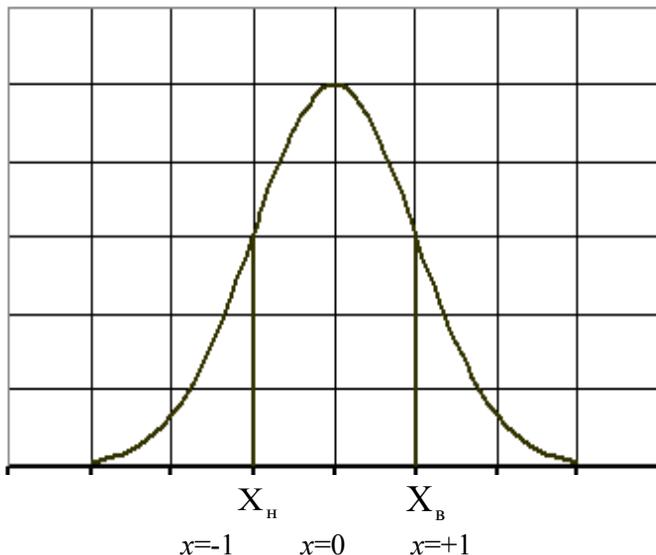


Рис. 1. Кодировка факторов в ММСБП

Очевидно, что для каждого фактора будут свои границы областей разбиения. При этом в таблицу исходных значений факторов, вместо самих значений факторов, подставляются их кодированные значения. Интересно отметить, что строки плана эксперимента, содержащие значения факторов равных 0, при дальнейшей обработке не используются, так, как алгоритм расчета коэффициентов при кодированных значениях факторов использует только области +1 и -1. Т.о. часть экспериментальных данных, принадлежащих центральной области нами не используется.

Назовем подобласти, на которые разбивают область определения каждого фактора опорными областями и введем следующие обозначения значений кодированного фактора: $x_1 = -1$, $x_2 = 0$, $x_3 = +1$, а средние значения выходного параметра Y при соответствующих значениях кодированного фактора обозначим как y_1 ,

y_2, y_3 . Результаты серии экспериментов, в разрезе одного конкретного фактора графически представлены на рис. 2.

Тогда описать такую зависимость можно уравнением 2-го порядка следующего вида: $y = ax^2 + bx + c$, где коэффициенты уравнения могут быть найдены из следующей системы уравнений:

$$(6) \quad \begin{cases} ax_1^2 + bx_1 + c = y_1 \\ ax_2^2 + bx_2 + c = y_2 \\ ax_3^2 + bx_3 + c = y_3 \end{cases}$$

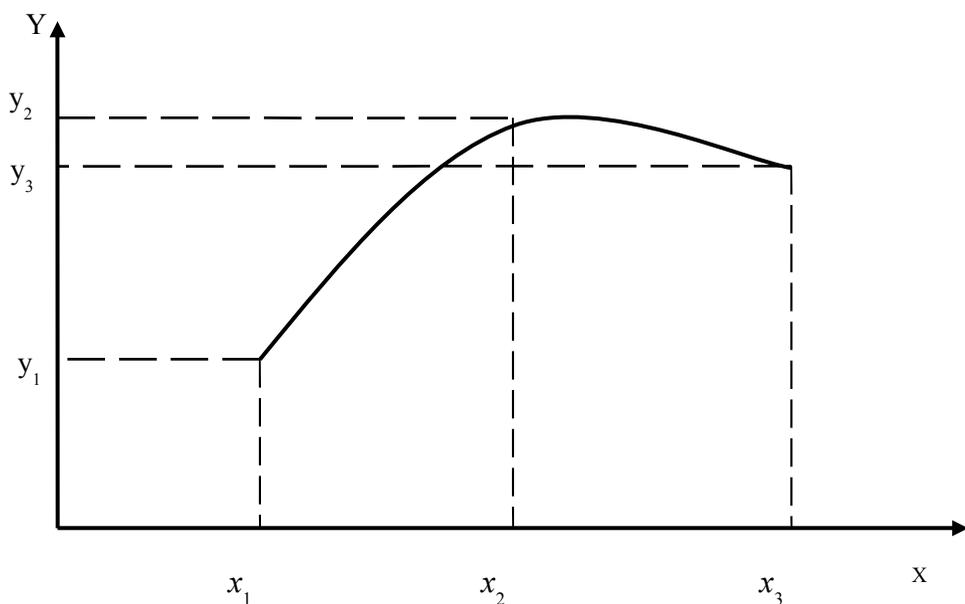


Рис. 2. Результаты эксперимента

Здесь приняты следующие обозначения: y_1, y_2, y_3 – это средние значения величины Y при $x=x_1, x=x_2, x=x_3$ соответственно. После решения этой системы мы получим следующие выражения для нахождения коэффициентов a, b, c :

$$(7) \quad c = y_2; \quad b = (y_3 - y_1)/2; \quad a = (y_3 + y_1)/2 - y_2.$$

Т.о., мы можем, используя все три опорные области, построить уравнение связи 2-го порядка, что в более ранних версиях ММСБП было невозможно.

Для проверки на значимость полученных коэффициентов модели используют выражение вида:

$$(8) \quad t_b = abs(b)/S_b < t_{таб}(q, \nu),$$

при выполнении которого коэффициент модели b признается незначимым (здесь ν - количество степеней свободы, q - уровень значимости, $t_{таб}(q, \nu)$ – распределение Стюдента, а S_b – стандартное отклонение коэффициента модели b).

Чтобы воспользоваться этим выражением, необходимо найти дисперсии коэффициентов модели с помощью выражения (9):

$$(9) \quad \begin{aligned} D[c] &= D[Y_2]/N_2; \\ D[b] &= (D[Y_3]/N_3 + D[Y_1]/N_1)/4; \\ D[a] &= (D[Y_3]/N_3 + D[Y_1]/N_1)/4 + D[Y_2]/N_2. \end{aligned}$$

Здесь использованы следующие обозначения:

$D[c]$, $D[b]$, $D[a]$ – дисперсии коэффициентов модели a , b , c ;
 $D[Y_2]$, $D[Y_3]$, $D[Y_1]$ – значения дисперсий выборки выходной величины Y для которых X принимает значения x_1 , x_2 , x_3 соответственно;

N_1 , N_2 , N_3 – объемы выборок выходной величины Y для которых X принимает значения x_1 , x_2 , x_3 соответственно.

Таким образом, предложенный нами метод может быть использован для получения моделей 2-го порядка по результатам пассивного эксперимента.

2.2 Численные результаты, полученные при построении модели 2-го порядка

Реализуем программно алгоритм, изложенный в 2.1. При этом границы опорных областей будем вычислять по формулам, представленным в таблице 1.

Построим математические модели для набора данных, взятых из [2, с.10]. Для сравнения будем строить следующие модели:

- линейную, содержащую только факторы в 1-й степени и не учитывающую их взаимодействия;
- квадратичную, с использованием формулы (7).

Таблица 1. Формулы, для определения границ опорных областей, для модели 2-го порядка

x_i	Границы областей
$x_1=-1$	$X < \bar{X} + k \cdot S_x$
$x_2=0$	$\bar{X} - k \cdot S_x < X < \bar{X} + k \cdot S_x$
$x_3=+1$	$X > \bar{X} + k \cdot S_x$

Здесь k - это некоторый коэффициент, задаваемый нами, который определяет размер центральной области, а следовательно, и границы остальных областей.

Для оценки качества модели воспользуемся индексом корреляции, который вычислим по формуле (10):

$$(10) \quad i = \sqrt{1 - \frac{\sum_{i=1}^N (Y_i - \hat{Y}_i)^2}{\sum_{i=1}^N (Y_i - \bar{Y})^2} * \frac{N-1}{N-m}} = \sqrt{1 - \frac{S_{\hat{Y}}^2}{S_Y^2}},$$

и суммой квадратов невязок, вычисляемых по формуле (11):

$$(11) \quad SS = \sum_{j=1}^N (Y_j - \hat{Y}_j)^2$$

где:

- N - это количество строк в таблице исходных данных;
- m - количество значимых коэффициентов модели;
- \hat{Y} - предсказанное с помощью модели значение величины Y .

Как известно, чем выше качество модели, тем ближе к единице величина индекса корреляции и тем ниже сумма квадратов невязок.

Используя разработанное нами программное обеспечение, проведем исследование поведения решения от величины опорных областей, определяемых коэффициентом k . Результаты вычислений приведены в таблице 2.

Для наглядности построим график по результатам, приведенным в таблице 2. Можно видеть, что качество модели очень сильно зависит от величины коэффициента k .

Таблица 2. Численные результаты, полученные при вычислениях, связанных с построением модели 2-го порядка

линейная модель								
k	SS	i	k	SS	i	k	SS	i
0,1	4725	0,753	0,7	4065	0,792	1,3	6178	0,659
0,2	4254	0,781	0,8	4039	0,794	1,4	6218	0,656
0,3	4229	0,783	0,9	4251	0,781	1,5	6429	0,641
0,4	4262	0,781	1	4705	0,754	1,6	6669	0,624
0,5	3995	0,796	1,1	5162	0,726	1,7	7266	0,578
0,6	3767	0,809	1,2	5581	0,699	1,8	8135	0,505
квадратичная модель								
k	SS	i	k	SS	i	k	SS	i
0,1	0,741	4919	0,7	0,67	6019	1,3	0,663	6119
0,2	0,584	7190	0,8	0,699	5577	1,4	0,676	5935
0,3	0,658	6186	0,9	0,695	5638	1,5	0,661	6156
0,4	0,698	5598	1	0,7	5568	1,6	0,647	6348
0,5	0,71	5433	1,1	0,712	5391	1,7	0,594	7065
0,6	0,719	5280	1,2	0,704	5515	1,8	0,512	8059

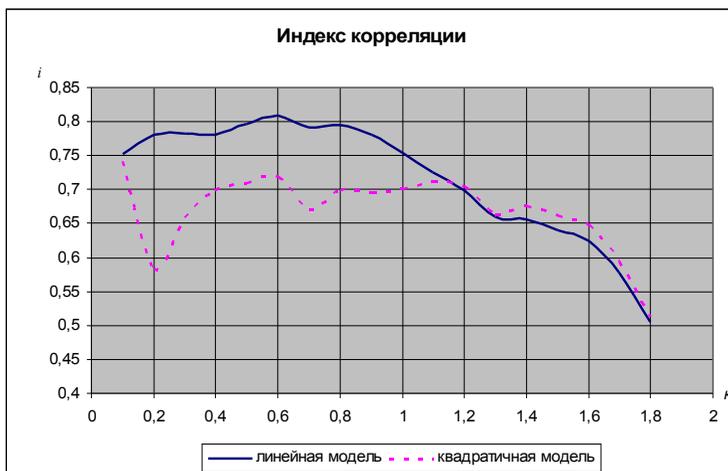


Рис. 3. Зависимость качества линейной и квадратичной моделей от величины опорных областей

Очевидно, что в данном случае предпочтительна линейная модель, что определяется реальным механизмом явлений, представленных в числовом материале. Квадратичная модель, начиная с $k=0,4$ демонстрирует стабильность, не изменяя существенно индекс корреляции до $k=1,2$.

2.3 Уравнение связи 3-го порядка

Рассмотрим случай, когда область определения факторов разбивается на 4 подобласти. Предлагаемое разбиение представлено на рис. 4:

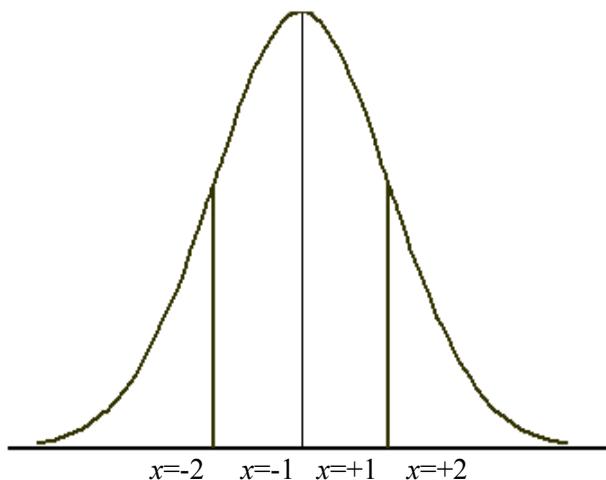


Рис. 4. Формирование 4-х опорных областей

Формирование границ областей представлено в таблице 3.

Таблица 3 - Границы 4-х опорных областей

Кодированное значение фактора	Границы опорной области
$x=-2$	$X < \bar{X} - kS_x$
$x=-1$	$\bar{X} - kS_x < X < \bar{X}$

$x = +1$	$\bar{X} < X < \bar{X} + kS_x$
$x = +2$	$X > \bar{X} + kS_x$

где k - это заданный нами коэффициент.

Обозначим возможные значения кодированного фактора: $x_1 = -2$, $x_2 = -1$, $x_3 = +1$, $x_4 = +2$. Результаты серии экспериментов, в разрезе одного конкретного фактора графически представлены на рис. 5.

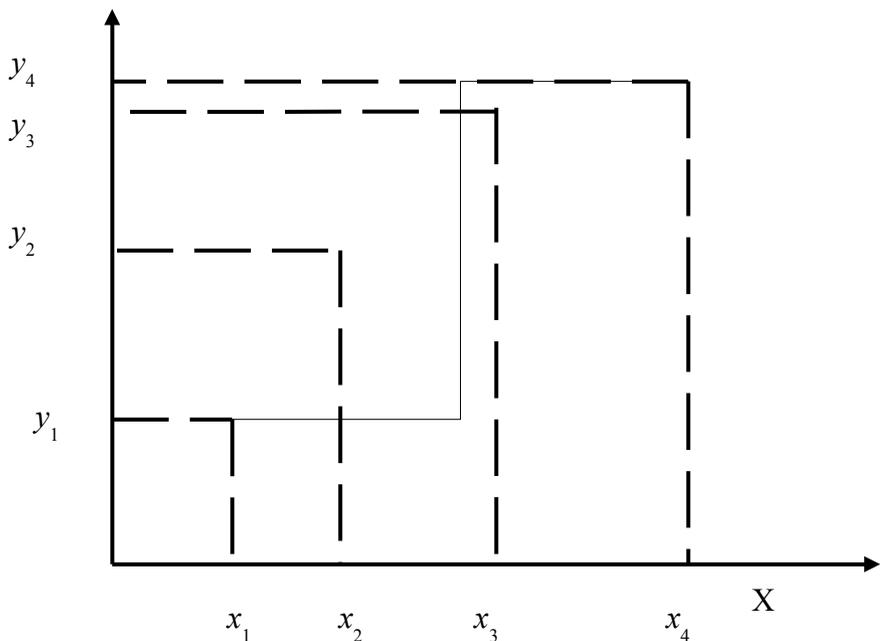


Рис. 5. Графическое представление результатов эксперимента

Здесь y_1, y_2, y_3, y_4 - это среднее значение выходной величины, для которых фактор принимает значения $-2, -1, +1, +2$ соответственно.

Кривая, изображенная на рис. 5, может быть описана уравнением вида: $y = ax^3 + bx^2 + cx + d$. Коэффициенты этого уравнения могут быть найдены из следующей системы уравнений:

$$(12) \quad \begin{cases} ax_1^3 + bx_1^2 + cx_1 + d = y_1 \\ ax_2^3 + bx_2^2 + cx_2 + d = y_2 \\ ax_3^3 + bx_3^2 + cx_3 + d = y_3 \\ ax_4^3 + bx_4^2 + cx_4 + d = y_4 \end{cases}$$

Подставив кодированные значения факторов в (12), и решив полученную систему, найдем:

$$(13) \quad \begin{aligned} a &= \frac{1}{12}(y_4 + 2y_2 - y_1 - 2y_3); \quad b = \frac{1}{6}(y_1 + y_4 - y_2 - y_3); \\ c &= \frac{1}{12}(8y_3 - y_4 + y_1 - 8y_2); \\ d &= \frac{1}{6}(4y_2 + 4y_3 - y_1 - y_4). \end{aligned}$$

Таким образом, мы получили коэффициенты уравнения модели третьего порядка.

Для оценки значимости коэффициентов модели воспользуемся выражением (8), причем дисперсии коэффициентов модели можно найти с помощью выражения (14).

$$(14) \quad \begin{aligned} D[a] &= (D[Y_1]/N_1 + 4D[Y_2]/N_2 + 4D[Y_3]/N_3 + D[Y_4]/N_4)/144; \\ D[b] &= (D[Y_1]/N_1 + D[Y_2]/N_2 + D[Y_3]/N_3 + D[Y_4]/N_4)/36; \\ D[c] &= (D[Y_1]/N_1 + 64D[Y_2]/N_2 + 64D[Y_3]/N_3 + D[Y_4]/N_4)/144; \\ D[d] &= (D[Y_1]/N_1 + 16D[Y_2]/N_2 + 16D[Y_3]/N_3 + D[Y_4]/N_4)/36. \end{aligned}$$

Здесь использованы следующие обозначения:

$D[a], D[b], D[c], D[d]$ – дисперсии коэффициентов модели a, b, c, d ;

$D[Y_1], D[Y_2], D[Y_3], D[Y_4]$ – значения дисперсий выборки выходной величины Y для которых X принимает значения x_1, x_2, x_3, x_4 соответственно;

N_1, N_2, N_3, N_4 – объемы выборок выходной величины Y для которых X принимает значения x_1, x_2, x_3, x_4 соответственно.

2.4 Численные результаты, полученные при построении модели 3-го порядка

Реализуем программно алгоритм, изложенный в 2.3. При этом границы опорных областей будем вычислять по формулам, представленным в таблице 3.

Построим математические модели для набора данных, взятых из [2, с.10]. Для сравнения будем строить следующие модели:

- линейную, содержащую только факторы в 1-й степени и не учитывающую их взаимодействия;
- кубичную, с использованием формул (13).

Для оценки качества модели воспользуемся индексом корреляции.

Используя разработанное нами программное обеспечение, проведем исследование поведения решения от величины опорных областей, определяемых коэффициентом k . Для этого будем менять величину коэффициента k и посмотрим, как это повлияет на качество модели. Результаты вычислений приведены в таблице 4.

Таблица 4. Численные результаты, полученные при вычислениях, связанных с построением модели 3-го порядка

Индекс корреляции i					
k	модель 3-го порядка	линейная модель	k	модель 3-го порядка	линейная модель
0,25	0,579	не адекватна	1,2	0,676	0,596
0,3	0,598	0,585	1,3	0,667	0,612
0,4	0,59	0,591	1,4	0,670	0,608
0,5	0,635	0,594	1,5	0,665	0,606
0,6	0,655	0,555	1,6	0,660	0,605
0,7	0,651	0,578	1,7	0,630	0,602
0,8	0,692	0,508	1,8	0,628	0,596
0,9	0,686	0,545	1,9	0,631	0,584
1	0,686	0,566	2	0,626	0,583
1,1	0,68	0,595			

График, построенный по данным, приведенным в таблице 4, представлен на рис. 6.

Можно видеть, что качество модели очень сильно зависит от величины коэффициента k , с другой стороны, при изменении k от 0,2 до 1, мы получаем модели с высоким значением индекса корреляции, который ведет себя достаточно стабильно.

Очевидно, что в данном случае модель третьего порядка является более точной, чем модель 1-го порядка.

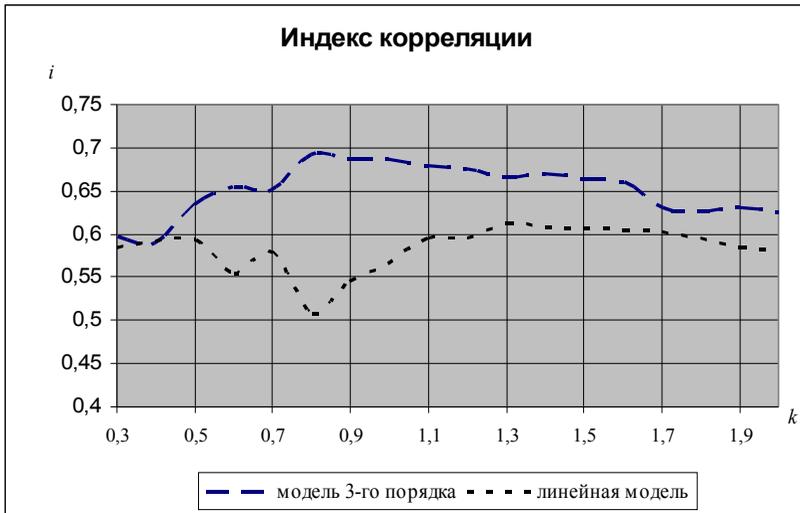


Рис. 6. Зависимость качества линейной модели и модели третьего порядка от величины опорных областей

Модель 3-го порядка, начиная с $k=0,7$ демонстрирует стабильность, не изменяя существенно индекс корреляции до $k=1,6$.

2.5 Уравнение связи 4-го порядка

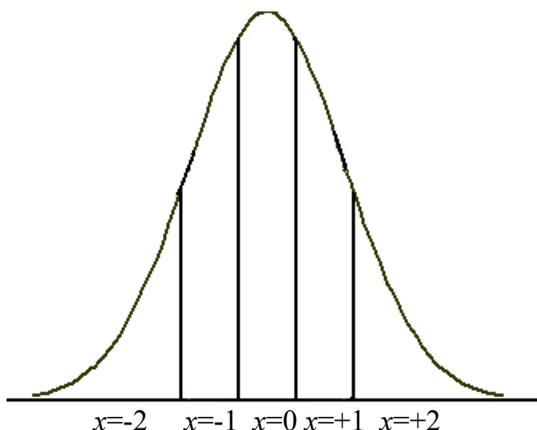


Рис. 7. Разбиение области определения фактора на 5 подобластей

Разбиение области определения величины X на 5 подобластей представлено в таблице 5.

Таблица 5 - Границы 5-и опорных областей

Кодированное значение фактора	Границы опорных областей
$x = -2$	$X < \bar{X} - 3kS_x$
$x = -1$	$\bar{X} - 3kS_x < X < \bar{X} - kS_x$
$x = -0$	$\bar{X} - kS_x < X < \bar{X} + kS_x$
$x = +1$	$\bar{X} + kS_x < X < \bar{X} + 3kS_x$
$x = +2$	$X > \bar{X} + 3kS_x$

где k - это заданный нами коэффициент.

В этом случае, после выполнения операции кодирования, наш фактор будет принимать следующие значения: -2, -1, 0, +1, +2. Введем следующие обозначения: $x_1=-2$, $x_2=-1$, $x_3=0$, $x_4=+1$, $x_5=+2$. Средние значения выходной величины Y для которых фактор принимает перечисленные выше значения, обозначим как y_1 , y_2 , y_3 , y_4 , y_5 . Результаты серии экспериментов, в разрезе

одного конкретного фактора графически представлены на рис. 8.

Кривую, представленную на рис. 8 можно описать уравнением вида: $y = ax^4 + bx^3 + cx^2 + dx + e$. Коэффициенты этого уравнения могут быть вычислены из следующей системы уравнений:

$$(15) \quad \begin{cases} ax_5^4 + bx_5^3 + cx_5^2 + dx_5 + e = y_5 \\ ax_4^4 + bx_4^3 + cx_4^2 + dx_4 + e = y_4 \\ ax_3^4 + bx_3^3 + cx_3^2 + dx_3 + e = y_3 \\ ax_2^4 + bx_2^3 + cx_2^2 + dx_2 + e = y_2 \\ ax_1^4 + bx_1^3 + cx_1^2 + dx_1 + e = y_1 \end{cases}$$

где приняты следующие обозначения: $x_1=-2$, $x_2=-1$, $x_3=0$, $x_4=+1$, $x_5=+2$, а y_1, y_2, y_3, y_4, y_5 – это средние значения выходной величины, которым соответствуют значения фактора $-2, -1, 0, +1, +2$.

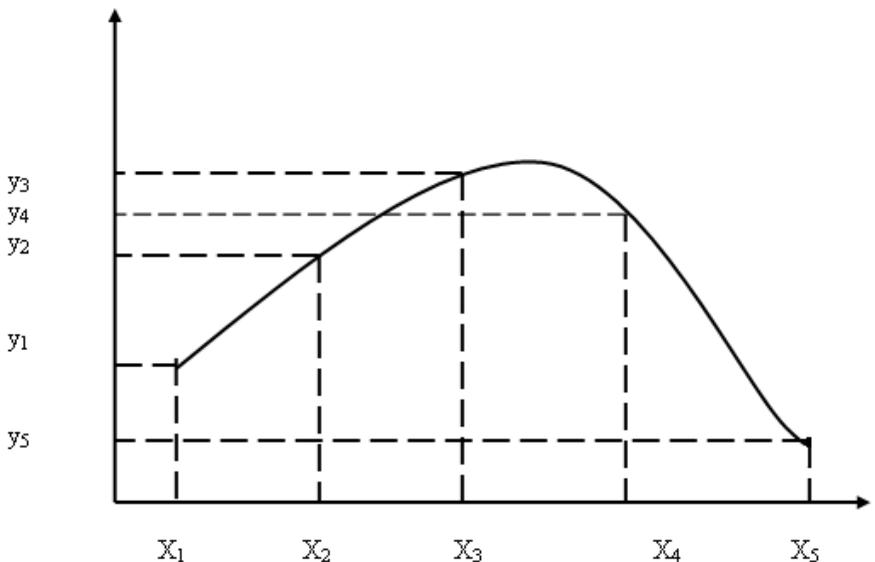


Рис. 8. Результаты эксперимента для 5-и опорных областей

Подставив в эту систему уравнений значения факторов и решив его, получим:

$$\begin{aligned}
 a &= \frac{1}{24}(6y_3 - 4y_4 - 4y_2 + y_1 + y_5); \\
 (16) \quad b &= \frac{1}{12}(y_5 - y_1 - 2y_4 + 2y_2); \\
 c &= \frac{1}{24}(16y_4 + 16y_2 - 30y_3 - y_5 - y_1); \\
 d &= \frac{1}{12}(y_1 - y_5 + 8y_4 - 8y_2); \quad e = y_3.
 \end{aligned}$$

Полученные выражения для коэффициентов a, b, c, d позволяют нам построить уравнение модели четвёртого порядка.

Для оценки значимости коэффициентов модели воспользуемся выражением (8), причем дисперсии коэффициентов модели можно найти с помощью выражения (17):

$$D[a] = (D[Y_1]/N_1 + 16D[Y_2]/N_2 + 36D[Y_3]/N_3 + 16D[Y_4]/N_4 + D[Y_5]/N_5)/576;$$

$$(17) \quad D[b] = (D[Y_1]/N_1 + 4D[Y_2]/N_2 + 4D[Y_4]/N_4 + D[Y_5]/N_5)/144;$$

$$D[c] = (D[Y_1]/N_1 + 256D[Y_2]/N_2 + 900D[Y_3]/N_3 + 256D[Y_4]/N_4 + D[Y_5]/N_5)/576$$

$$D[d] = (D[Y_1]/N_1 + 64D[Y_2]/N_2 + 64D[Y_4]/N_4 + D[Y_5]/N_5)/144.$$

Здесь использованы следующие обозначения:

$D[a], D[b], D[c], D[d]$ – дисперсии коэффициентов модели a, b, c, d ;

$D[Y_1], D[Y_2], D[Y_3], D[Y_4], D[Y_5]$ – значения дисперсий выборки выходной величины Y для которых X принимает значения x_1, x_2, x_3, x_4, x_5 соответственно;

N_1, N_2, N_3, N_4, N_5 – объемы выборок выходной величины Y для которых X принимает значения x_1, x_2, x_3, x_4, x_5 соответственно.

2.6 Численные результаты, полученные при построении модели 4-го порядка

Реализуем программно алгоритм, изложенный в 2.5. При этом границы опорных областей будем вычислять по формулам, представленным в таблице 5.

Построим математические модели для набора данных, взятых из [2, с.10]. Для сравнения будем строить следующие модели:

- линейную, содержащую только факторы в 1-й степени и не учитывающую их взаимодействия;

- кубичную, с использованием формул (16).

Для оценки качества модели воспользуемся индексом корреляции.

Используя разработанное нами программное обеспечение, проведем исследование поведения решения от величины опорных областей, определяемых коэффициентом k . Для этого будем менять величину коэффициента k и посмотрим, как это повлияет на качество модели. Результаты вычислений приведены в таблице 6.

График, построенный по данным, приведенным в таблице 6 представлен на рис. 9. Можно видеть, что качество модели очень сильно зависит от величины коэффициента k , с другой стороны, при изменении k от 0,2 до 1, мы получаем модели с высоким значением индекса корреляции, который ведет себя достаточно стабильно.

Таблица 6. Численные результаты, полученные при вычислениях, связанных с построением модели 3-го порядка

Индекс корреляции i					
k	модель 3-го порядка	линейная модель	k	модель 3-го порядка	линейная модель
0,1	0,385	не адекватна	1,1	0,68	0,595
0,2	0,588		1,2	0,676	0,596
0,25	0,579		1,3	0,667	0,612
0,3	0,598	0,585	1,4	0,670	0,608
0,4	0,59	0,591	1,5	0,665	0,606
0,5	0,635	0,594	1,6	0,660	0,605
0,6	0,655	0,555	1,7	0,630	0,602
0,7	0,651	0,578	1,8	0,628	0,596
0,8	0,692	0,508	1,9	0,631	0,584
0,9	0,686	0,545	2	0,626	0,583
1	0,686	0,566			



Рис. 9. Зависимость качества линейной модели и модели третьего порядка от величины опорных областей

Очевидно, что в данном случае модель третьего порядка является более точной, чем модель 1-го порядка. Наибольшая точность модели достигается при $0,7 < k < 1,6$. Модель 3-го порядка, начиная с $k=0,7$ демонстрирует стабильность, не изменяя существенно индекс корреляции до $k=1,6$.

2.7 Построение уравнения связи 2-го порядка в условиях гетероскедастичности

Будем искать уравнение связи в следующем виде: $y = ax^2 + bx + c$. Рассмотрим коэффициент при x^2 , который вычисляется по формуле: $a = (y_3 + y_1)/2 - y_2$, и введем весовые коэффициенты $\beta_1, \beta_2, \beta_3$, причем $\beta_1 + \beta_2 + \beta_3 = 1$, тогда:

$$(18) \quad a = \beta_3 y_3 / 2 + \beta_1 y_1 / 2 - \beta_2 y_2,$$

где $\beta_2 = 1 - \beta_1 - \beta_3$.

Введем следующие обозначения: $D_1 = D[Y_1]$, $D_2 = D[Y_2]$, $D_3 = D[Y_3]$ – значения дисперсий выборки выходной величины Y для которых X принимает значения x_1, x_2, x_3 соответственно;

N_1, N_2, N_3 – объемы выборок выходной величины Y для которых X принимает значения x_1, x_2, x_3 соответственно;

$$\frac{D_3}{N_3} = D^3; \frac{D_2}{N_2} = D^2; \frac{D_1}{N_1} = D^1 \quad \text{и} \quad \frac{D^1}{D^2} = D^{1/2}; \frac{D^3}{D^2} = D^{3/2}.$$

Записав выражения для дисперсии коэффициента а нашей модели и решив задачу поиска коэффициентов $\beta_1, \beta_2, \beta_3$, при которых данная дисперсия будет минимальна, получим следующие выражения:

$$(19) \quad \beta_1 = 1/(0.25D^{1/2} + D^{1/2}/D^{3/2} + 1);$$

$$(20) \quad \beta_3 = 1/(0.25D^{3/2} + D^{3/2}/D^{1/2} + 1);$$

$$(21) \quad \beta_2 = 1 - \beta_1 - \beta_3.$$

Полученные нами $\beta_1, \beta_2, \beta_3$ - весовые коэффициенты, необходимые для вычисления коэффициента при x_2 , обеспечивающие его минимальную дисперсию.

Оптимальная оценка для коэффициента модели b получена ранее в [1] и имеет вид:

$$(22) \quad b = \left((D^1 + 2\bar{y}^{-2})y_3 - (D^3 + 2\bar{y}^{-2})y_1 \right) / (D^3 + D^1 + 4\bar{y}^{-2})$$

Таким образом, мы получили выражения, позволяющие построить в условиях гетероскедастичности модель второго порядка, коэффициенты которой имеют минимально возможную дисперсию.

2.8 Численные результаты, полученные при построении модели 2-го порядка в условиях гетероскедастичности

Реализуем программно алгоритм, изложенный в 2.7. При этом границы опорных областей будем вычислять по формулам, представленным в таблице 7.

Таблица 7. Формулы, для определения границ опорных областей, для модели 2-го порядка

x_i	Границы областей
$x_1 = -1$	$X < \bar{X} + k \cdot S_x$
$x_2 = 0$	$\bar{X} - k \cdot S_x < X < \bar{X} + k \cdot S_x$
$x_3 = +1$	$X > \bar{X} + k \cdot S_x$

Здесь k - это некоторый коэффициент, задаваемый нами, который определяет размер центральной области, а следовательно, и границы остальных областей.

Построим математические модели для набора данных, взятых из [2, с.10]. Для сравнения будем строить следующие модели:

- линейную, содержащую только факторы в 1-й степени и не учитывающую их взаимодействия;
- квадратичную, с использованием формул (7);
- квадратичную, с учетом гетероскедастичности по формуле (18).

Для оценки качества модели воспользуемся индексом корреляции.

Будем менять величину коэффициента k и посмотрим, как это повлияет на качество модели. Результаты вычислений приведены в таблице 8.

Как видно из таблицы 8 при значениях $k \geq 0.5$ модель, построенная с учетом гетероскедастичности значительно точнее прочих моделей.

Таблица 8. Численные результаты, полученные при вычислениях, связанных с построением модели 2-го порядка

k	Значения индекса корреляции i для различных видов модели		
	Линейная модель	Квадратичная модель	Квадратичная модель с учетом гетероскедастичности
0,1	0,753	0,741	0,546
0,2	0,781	0,584	0,712
0,3	0,783	0,658	0,756
0,4	0,781	0,698	0,781
0,5	0,796	0,71	0,885
0,6	0,809	0,719	0,926
0,7	0,792	0,67	0,953

7			
0,8	0,794	0,699	0,964
0,9	0,781	0,695	0,976
1,0	0,754	0,7	0,981
1,1	0,726	0,712	0,985
1,2	0,699	0,704	0,985
1,3	0,659	0,663	0,985
1,4	0,656	0,676	0,983
1,5	0,641	0,661	0,981
1,6	0,624	0,647	0,980

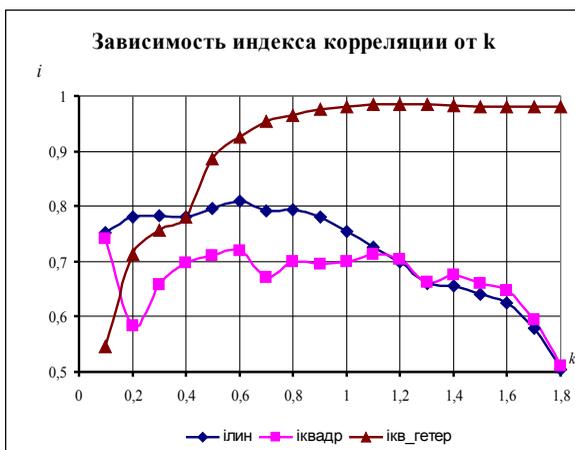


Рис. 10. Зависимость качества моделей от размера центральной опорной области.

Чтобы увидеть это более наглядно, построим по содержанию таблицы 8 графики, представленные на рис. 10. Видно, что

начиная с $k=0.5$ квадратичная модель, найденная с учетом гетероскедастичности, предпочтительнее всех конкурирующих моделей.

3. Выводы

1. Проведенная работа позволила получить выражения, позволяющие построить по данным пассивного эксперимента модель высокого порядка с помощью Модифицированного метода случайного баланса для пассивного эксперимента (ММСБП).

2. Получены и апробированы выражения, позволяющие построить модель второго порядка в условиях гетероскедастичности.

3. Показано, что для модели второго порядка учет гетероскедастичности резко повышает ее точность.

4. Возможна дальнейшая модификация ММСБ для моделей более высоких порядков.

Литература

1. Долгов Ю. А. Модифицированный метод случайного баланса // Электрон. моделирование.-1987.-9, № 4.- с. 79 - 84.
2. Долгов Ю.А. Статистическое моделирование. – Тирасполь: РИО ПГУ, 2002. -280 с.
3. Федорченко С.Г., Данейкин А.А. Моделирование высокого порядка по результатам пассивного эксперимента// Вестник ПГУ, сер. физ.-мат. и технич. наук. – 2007.- N 3. - С.117-122.
4. Федорченко С.Г. Построение модели по результатам пассивного эксперимента//Радиоэлектроника, информатика, управление. Запорожский государственный технический университет, № 1, 1999.-с.92-95.

Build the mathematical models on results a passive experiment

Sergey Fedorchenko, State University, Tiraspol, Doctor of Science
(fed_tir@mail.ru).

Abstract

Development of the modified of causal balance? Allowing to build the mathematical models of high order on results a passiv experiment in considered in the article

Passive experiment? Normal distribution? Factor orthogonality?
Support areas.