УДК 519.233.5

С.Г. РАДЧЕНКО*

СТАТУС МАТЕМАТИЧЕСКИХ МОДЕЛЕЙ, ПОЛУЧАЕМЫХ С ИСПОЛЬЗОВАНИЕМ РЕГРЕССИОННОГО АНАЛИЗА

*Национальный технический университет Украины «Киевский политехнический институт», Киев, Украина

Анотація. *Проаналізовано умови одержання статистичних моделей при використанні регресійного аналізу. Показано, що задачі, які вирішуються, є некоректно поставленими. Викладено методи рішення задач – планування експерименту та регуляризація. Розглянуто приклад одержання моделей у багатофакторних і багатокритеріальних задачах.*
Ключові слова: *планування експерименту, регуляризація, регресійний аналіз, некоректно поставлені задачі.*

Аннотация. *Проанализированы условия получения статистических моделей при использовании регрессионного анализа. Показано, что решаемые задачи являются некорректно поставленными. Изложены методы решения задач – планирование эксперимента и регуляризация. Рассмотрен пример получения моделей в многофакторных и многокритериальных задачах.*
Ключевые слова: *планирование эксперимента, регуляризация, регрессионный анализ, некорректно поставленные задачи.*

Abstract. *Conditions for obtaining statistical models using regression analysis have been analyzed. It is shown that the problems for solving are incorrectly posed. The methods of problems solution have been stated: the experiment design and regularization. An example of obtaining models in multifactor and multi-criterion problems has been considered.*
Keywords: *experiment design, regularization, regression analysis, incorrectly posed problems.*

1. Введение. Постановка проблемы

Создание и совершенствование наукоемких изделий, высоких технологий, интеллектуальных средств измерений возможно при использовании информационных технологий в виде многокритериальной компромиссной оптимизации и многофакторного математического моделирования. Новизна и сложность систем и процессов затрудняют использование теоретико-аналитического подхода.

Исходные данные для получения моделей часто характеризуются отсутствием необходимой информации: структуры многофакторной модели, статистически значимых факторов, закона распределения моделируемого критерия и др. Полученные данные представляют суммарное влияние групп управляемых, неуправляемых и неконтролируемых факторов.

Решаемая задача получения моделей относится к классу обратных задач: по полученным данным восстановить влияние факторов в виде главных эффектов и взаимодействий эффектов.

Исходные данные содержат случайные и систематические ошибки результатов экспериментов, измерений, вычислений. Задача получения моделей, будучи обратной, в большинстве случаев является некорректно поставленной. Возможно получение различных по структуре моделей, малые исходные погрешности могут приводить к сравнительно большим изменениям конечных получаемых результатов.

Цель статьи

Определить статус получаемой модели и сформулировать условия построения многофакторных статистических моделей в регрессионном анализе, которые позволяют решать корректно поставленные задачи.

2. Анализ достижений и публикаций по теме исследования

При использовании математического аппарата для изучения реальных сложных систем необходимо учитывать его принципиальные свойства и вытекающие из них возможности. Исследование математических структур в классической математике происходит на основе закона классической логики – закона исключенного третьего: одно из двух высказываний "A" и "не A" является истинным, третьего не дано [1, с. 96]. В математической логике это соглашение задается аксиомой $A \vee \rceil A$. Тогда результатом логического вывода могут быть высказывания типа «да» и «нет». Высказывания типа «может быть», «возможно при некоторых формально не заданных условиях» не рассматриваются. Для применяемого математического аппарата необходима исчерпывающая исходная информация о свойствах и характеристиках изучаемых сложных систем. Обеспечить полноту и истинность исходной и последующей необходимой информации об их свойствах не представляется возможным.

Акад. И.И. Артоболевский рассматривал технологические системы как системы с неполной информацией [2, с. 16]. «Процессу функционирования больших технологических систем и процессу их синтеза свойственна известная неопределенность, вызванная неполнотой информации об условиях эксплуатации, о качестве используемых систем и т.п.». Отмечается, что вследствие множества варьируемых переменных и стохастического характера этих переменных, начальных условий, проектирование сложных технологических систем машин и линий встречает большие математические трудности [2, с. 17].

Проф. В.А. Вознесенский также обращал внимание на указанные свойства сложных систем. «К сожалению, в технологических, технико-экономических и других реальных сложных системах неизвестны ни вид функции φ (уравнение состояния системы – Р.С.), ни полный набор факторов $\mathbf{X}_{\varphi}$, ни числовые значения констант (коэффициентов) $\mathbf{B}_{\varphi}$, ни законы распределения случайных величин ε_{φ}, ни граничные условия» [3, с. 12].

Организация эксперимента в экономических, социальных условиях, процессах, происходящих в естественной природе, затруднена или невозможна. Результаты происходящих процессов фиксируются как наблюдения, а факторы являются неуправляемыми. Для таких процессов типичным является коррелируемость факторов и, следовательно, невозможность раздельно управлять каждым фактором по определенной схеме.

Типичность плохой обусловленности прикладных задач самых разнообразных классов стимулировала поиск математиками и специалистами в прикладных областях исследований общих методов их решения. Получил определенное распространение метод регуляризации некорректно поставленных задач в различных модификациях. Значительные работы были проведены школами акад. А.Н. Тихонова, акад. М.М. Лаврентьева и их учениками и последователями [4; 5, с. 161–171; 6, с. 47–72].

Известны также методы статистической регуляризации В.Ф. Турчина, Ю.Е. Воскобойникова, итеративной регуляризации В.М. Фридмана, робастной регуляризации В.Я. Арсенина. Разработанные методы позволили решать неустойчивые прикладные задачи обработки сигналов и изображений в области астрофизики, томографии, оптики, спектроскопии, ядерной физики и в других областях, в математических моделях которых число обусловленности $\mathrm{cond} \approx 10^4 \ldots 10^6$, а после регуляризации становилось $\mathrm{cond} \approx 10^2$.

Получаемые с использованием регрессионного анализа модели являются в общем случае некорректно поставленными задачами. Отсутствие информации о структуре модели и наличие ошибок при получении коэффициентов модели требуют использования специальных устойчивых методов решения задачи.

3. Решение проблемы

Будем рассматривать получение полиномиальной модели линейной по параметрам и нелинейной по факторам. Использование модели базируется на теореме Вейерштрасса о возможности как угодно близкой аппроксимации произвольной действительной непрерывной на отрезке $[a, b]$ функции $f(x)$ последовательностью алгебраических многочленов $P_0(x), P_1(x), ..., P_n(x), ...,$ равномерно сходящейся на $[a, b]$ к функции $f(x)$, и на теоремах М. Стоуна и Д. Джексона.

Приведенная неопределенность состояния сложных систем является весьма типичной и затрудняет определенно решать поставленную задачу получения математических моделей.

Статус получаемой многофакторной статистической модели классифицируется как некорректно поставленная задача. Устранение неопределенности при получении модели возможно, если использовать методы теории планирования эксперимента. Для этого необходимо, чтобы была возможность организации эксперимента по определенной схеме. Как правило, это возможно в технических, естественно-научных экспериментах, которые можно поставить в лабораторных или промышленных условиях.

В качестве плана эксперимента будем использовать полный факторный эксперимент. По теореме В.З. Бродского, все полиномиальные эффекты и их взаимодействия ортогональны друг к другу.

Формализованная структура многофакторной статистической модели задается выражением

$$\prod_{i=1}^{k}(1 + x_i^{(1)} + x_i^{(2)} + ... + x_i^{(s_i-1)}) \to N_\Pi ,$$

где 1 – значение фиктивного фактора $x_0 \equiv 1$;

$x_i^{(1)},\ x_i^{(2)}, ...,\ x_i^{(s_i-1)}$ – ортогональные контрасты факторов X_i;

s_i – число различных уровней фактора X_i;

k – общее число факторов, $1 \le i \le k$;

(1), (2), …, $(s_i - 1)$ – порядок контрастов фактора X_i;

N_Π – число структурных элементов полного факторного эксперимента, равное числу опытов эксперимента.

При использовании дробных факторных экспериментов применяют многофакторные регулярные планы и планы на основе $ЛП_\tau$ равномерно распределенных последовательностей.

Получаемые статистические модели проверяются по следующим критериям [7, с. 65–80].

1. Проверка на статистическую значимость модели.
2. Проверка коэффициентов уравнения регрессии на статистическую значимость.
3. Проверка на адекватность
4.Проверка на информативность.
5. Проверка на устойчивость.
6. Проверка фактической эффективности извлечения полезной информации.
7. Проверка правильности описания полученной модели по всей области моделирования.
8. Оценка семантичности информационной.
9. Проверка свойств остатков.
10. Общая оценка свойств полученной математической модели.

Для получения модели используют алгоритм RASTA3 и программное средство «Планирование, регрессия и анализ моделей» (ПС ПРИАМ) [7, с. 81–85]. Неизвестная структура многофакторной модели выбирается из множества структурных элементов схемы полного факторного эксперимента.

Изложенная методология позволяет получить модели, наиболее близкие к «истинным» моделям [7].

Основным требованием при получении статистических моделей является ортогональность плана эксперимента. Расширенная концепция ортогональности и нормирование эффектов – главных и взаимодействий – позволяет устранить неопределенность в выборе структуры модели и ее коэффициентов. Число обусловленности полученной модели cond не превышает 10. Ограничение в использовании концепции ортогональности связано с числом проведенных опытов. План эксперимента не должен быть близким к насыщенному.

Отметим, что обеспечение концепции ортогональности связано с выбором системы координат особенно в тех случаях, когда форма факторного пространства отличается от формы прямоугольного параллелепипеда [7, с. 134–186].

Нестандартная область факторного пространства рассматривается как область образа, которому топологически соответствует область прообраза. Между образом и прообразом устанавливается взаимно однозначное и взаимно непрерывное соответствие. В области прообраза эксперимент можно планировать и получать математические модели с наилучшими статистическими свойствами. Отображая прообраз в образ, получаем информацию в образе, соответствующую наилучшим возможным свойствам: статистическая эффективность, устойчивость.

Рассмотрим пример решения реальной задачи многофакторного математического моделирования модульной сборки многоэлементных конструкций. В качестве исходной информации используются результаты метода имитационного моделирования на ЭВМ процесса сборки многоэлементных конструкций [8].

Для получения моделей был выбран равномерный план эксперимента 4^7 //32.

Исследование закономерностей образования погрешности подсистемы «Подсоединяемая деталь» производилось для критериев качества (откликов) $y_4 ... y_9$ (табл. 1), изучалось влияние на них параметров $X_1 ... X_7$ (табл. 2).

Таблица 1. Критерии качества системы

Наименование функции отклика	Обозначения критериев качества	
	натуральные	формализованные
Главные средние квадратические отклонения оси поверхности сопряжения присоединяемой детали (центрирующее захватное устройство)	PSH	y_4
	PST	y_5
Среднее квадратическое отклонение модуля погрешности оси поверхности сопряжения присоединяемой детали (центрирующее захватное устройство)	PDM	y_6
Главные средние квадратические отклонения оси поверхности сопряжения присоединяемой детали (призматическое захватное устройство)	$PSH1$	y_7
	$PST1$	y_8
Среднее квадратическое отклонение модуля погрешности оси поверхности сопряжения присоединяемой детали (призматическое захватное устройство)	$PDM1$	y_9

Таблица 2. Исследуемые факторы подсистемы «Присоединяемая деталь»

Наименование фактора	Обозначения факторов		Значения уровней факторов F_i			
			0	1	2	3
	натуральные	формализованные	Значения факторов X_i на уровнях s_i			
Линейная погрешность позиционирования промышленного робота, мм	SPX	X_1	0,01	0,18	0,35	0,52
Угловая погрешность позиционирования промышленного робота, рад	$GP2$	X_2	0,002	0,008	0,014	0,020
Допуск на взаимное расположение поверхности базирования в зажимном устройстве и поверхности сопряжения присоединяемой детали, мм	$DS2$	X_3	0,04	0,16	0,28	0,40
Относительная длина поверхности сопряжения присоединяемой детали	DCM / DLD	X_4	0,05	0,20	0,35	0,50
Длина присоединяемой детали от торца поверхности сопряжения до центра зажимного устройства, мм	$DL2$	X_5	20	100	180	260
Допуск на диаметр базовой поверхности присоединяемой детали при базировании в зажимном устройстве, мм	$DB2$	X_6	0,010	0,034	0,067	0,100
Угол ориентации конечного звена промышленного робота при установке присоединяемой детали, рад	BE	X_7	0	1,05	2,09	3,14

Рабочая матрица планирования, уровни варьирования факторов и результаты опытов второй серии экспериментов приведены в табл. 3.

Так как исходные данные для всех моделей были получены методом имитационного моделирования, то дублирование опытов не проводилось и дисперсия воспроизводимости определялась как квадрат среднего квадратического отклонения исследуемого отклика. Число степеней свободы было условно принято 32.

Параметры закона распределения определялись обработкой результатов эксперимента на ПС ПРИАМ.

Для формирования структур моделей $\hat{y}_4(PSH) \ldots \hat{y}_9(PDM1)$ были проанализированы 21 главный эффект и 189 эффектов двойных взаимодействий, то есть всего 210 эффектов. При выборе эффектов кандидатов для ввода в модели были использованы следующие условия:

– ограничение максимального коэффициента корреляции между эффектами 0,4;

– ограничение минимального коэффициента корреляции эффектов с откликом 0,01;

– ограничение минимальной доли рассеяния 0,005.

Список эффектов кандидатов для возможного использования в математических моделях включал 31 главный эффект и двойные взаимодействия.

Таблица 3. Уровни варьирования, рабочая матрица, результаты опытов

Номер опыта	Значения уровней факторов								Значения функций отклика					
	F_i	X_1	X_2	X_3	X_4	X_5	X_6	X_7	y_4	y_5	y_6	y_7	y_8	y_9
	0	0,01	0,002	0,04	0,05	20	0,010	0						
	1	0,18	0,008	0,16	0,20	100	0,034	1,05						
	2	0,35	0,014	0,28	0,35	180	0,067	2,09						
	3	0,52	0,020	0,40	0,50	260	0,100	3,14						
1		0,01	0,002	0,04	0,05	20	0,010	0	0,02	0,02	0,01	0,02	0,02	0,11
2		0,01	0,002	0,28	0,50	180	0,100	1,05	0,11	0,12	0,077	0,11	0,12	0,08
3		0,01	0,008	0,04	0,50	100	0,067	3,14	0,17	0,18	0,11	0,17	0,18	0,11
4		0,01	0,008	0,28	0,05	260	0,034	2,09	0,56	0,61	0,39	0,56	0,61	0,39
5		0,01	0,014	0,16	0,20	260	0,100	0	0,93	0,97	0,62	0,93	0,97	0,61
6		0,01	0,014	0,40	0,35	100	0,010	1,05	0,34	0,37	0,24	0,33	0,37	0,24
7		0,01	0,020	0,16	0,35	180	0,034	3,14	0,85	0,88	0,57	0,87	0,88	0,55
8		0,01	0,020	0,40	0,20	20	0,067	2,09	0,15	0,16	0,11	0,15	0,16	0,11
9		0,18	0,008	0,16	0,20	100	0,034	1,05	0,22	0,23	0,15	0,22	0,23	0,15
10		0,18	0,008	0,40	0,35	260	0,067	0	0,49	0,53	0,34	0,49	0,53	0,34
11		0,18	0,002	0,16	0,35	20	0,100	2,09	0,07	0,07	0,05	0,08	0,07	0,05
12		0,18	0,002	0,40	0,20	180	0,010	3,14	0,15	0,16	0,11	0,15	0,16	0,11
13		0,18	0,020	0,04	0,05	180	0,067	1,05	1,00	1,03	0,67	1,00	1,04	0,65
14		0,18	0,020	0,28	0,50	20	0,034	0	0,13	0,14	0,09	0,13	0,14	0,09
15		0,18	0,014	0,04	0,50	260	0,010	2,09	0,78	0,81	0,52	0,78	0,81	0,50
16		0,18	0,014	0,28	0,05	100	0,100	3,14	0,41	0,42	0,27	0,41	0,42	0,27
17		0,35	0,014	0,28	0,35	180	0,067	2,09	0,61	0,63	0,41	0,62	0,63	0,40
18		0,35	0,014	0,04	0,20	20	0,034	3,14	0,14	0,13	0,09	0,14	0,13	0,09
19		0,35	0,020	0,28	0,20	260	0,010	1,05	1,34	1,39	0,90	1,34	1,39	0,90
20		0,35	0,020	0,04	0,35	100	0,100	0	0,49	0,50	0,32	0,49	0,50	0,31
21		0,35	0,002	0,40	0,50	100	0,034	2,09	0,17	0,17	0,11	0,17	0,17	0,11
22		0,35	0,002	0,16	0,05	260	0,067	3,14	0,20	0,20	0,13	0,20	0,20	0,13
23		0,35	0,008	0,40	0,05	20	0,100	1,05	0,17	0,17	0,11	0,17	0,17	0,11
24		0,35	0,008	0,16	0,50	180	0,010	0	0,34	0,34	0,22	0,34	0,34	0,22
25		0,52	0,020	0,40	0,50	260	0,100	3,14	1,14	1,18	0,76	1,14	1,18	0,74
26		0,52	0,020	0,16	0,05	100	0,010	2,09	0,59	0,60	0,39	0,59	0,60	0,38
27		0,52	0,014	0,40	0,05	180	0,034	0	0,74	0,76	0,49	0,74	0,76	0,48
28		0,52	0,014	0,16	0,50	20	0,067	1,05	0,19	0,18	0,12	0,19	0,18	0,12
29		0,52	0,008	0,28	0,35	20	0,010	3,14	0,19	0,19	0,13	0,19	0,19	0,12
30		0,52	0,008	0,04	0,20	180	0,100	2,09	0,41	0,42	0,27	0,41	0,42	0,27
31		0,52	0,002	0,28	0,20	100	0,067	0	0,20	0,19	0,13	0,20	0,19	0,13
32		0,52	0,002	0,04	0,35	260	0,034	1,05	0,21	0,21	0,14	0,22	0,20	0,14

В результате получены следующие математические модели.

Для центрирующего зажимного устройства

Главные средние квадратические отклонения оси поверхности сопряжения присоединяемой детали в плоскости сборки:

$$\hat{y}_4(PSH) = 0{,}422188 + 0{,}288563\,x_5 + 0{,}286313\,x_2 + 0{,}246038\,x_2 x_5 -$$

$$-0{,}0425625\,x_4 + 0{,}0343125\,x_1,$$

$$\hat{y}_5(PST) = 0{,}43625 + 0{,}30375\,x_5 + 0{,}296625\,x_2 + 0{,}253575\,x_2 x_5 - 0{,}043875\,x_4,$$

где $x_1 = 3{,}92157(X_1 - 0{,}265)$, $x_2 = 111{,}111(X_2 - 0{,}011)$,
$x_4 = 4{,}44444(X_4 - 0{,}275)$, $x_5 = 0{,}00833333(X_5 - 140)$.

В приведенных формулах ортогональные контрасты предварительно не нормируются. Нормирующие множители $k_i^{(p)}$ учтены в коэффициентах полученных математических моделей.

Для модели $\hat{y}_6(PDM)$ вместо основного алгоритма выделения структуры модели был использован алгоритм последовательного выделения структуры. Он основан на вычислении частных коэффициентов корреляции между эффектами и откликом.

Среднее квадратическое отклонение модуля погрешности расположения поверхности сопряжения присоединяемой детали:

$$\hat{y}_6(PDM) = 0{,}282719 + 0{,}194381\,x_5 + 0{,}191231\,x_2 + 0{,}173366\,x_2 x_5 -$$

$$-0{,}0288563\,x_4 - 0{,}0327408\,x_4 x_5.$$

Для зажимного устройства с базированием присоединяемой детали в призме.

Главные средние квадратические отклонения:

$$\hat{y}_7(PSH1) = 0{,}423438 + 0{,}289313\,x_5 + 0{,}286313\,x_2 + 0{,}246938\,x_2 x_5 -$$

$$-0{,}0418125\,x_4 + 0{,}0343125\,x_1,$$

$$\hat{y}_8(PST1) = 0{,}43625 + 0{,}303375\,x_5 + 0{,}29775\,x_2 + 0{,}254925\,x_2 x_5 - 0{,}044625\,x_4.$$

Среднее квадратическое отклонение модуля погрешности расположения оси поверхности сопряжения присоединяемой детали:

$$\hat{y}_9(PDM1) = 0{,}281562 + 0{,}185812\,x_5 + 0{,}180188\,x_2 + 0{,}168638\,x_2 x_5 - 0{,}0350625\,x_4.$$

Все критерии качества полученных математических моделей (табл. 4) являются наилучшими: модели адекватны, высоко информативны, устойчивы. В модели $\hat{y}_6(PDM)$ взаимодействие $x_4 x_5$ коррелировано с взаимодействием $x_2 x_5$ с коэффициентом парной корреляции 0,32. Все остальные эффекты во всех моделях ортогональны друг к другу. Средние абсолютных величин погрешностей аппроксимации $|\bar{e}_u|$ сравнительно малы. Доли рассеивания, объясняемые моделями, достаточно близки к 1 (все превышают 0,98).

В полученных моделях все эффекты ортогональны друг к другу и нормированы (слабокоррелированы только два эффекта в модели $\hat{y}_6(PDM)$. Поэтому полученные математические модели можно использовать для установления причинных, структурных и количественных связей между комплексом исходных условий и группой моделируемых критериев качества.

Таблица 4. Результаты статистического анализа математических моделей $\hat{y}_w$

Параметры статистического анализа		Значения параметров для моделей					
		$\hat{y}_4$	$\hat{y}_5$	$\hat{y}_6$	$\hat{y}_7$	$\hat{y}_8$	$\hat{y}_9$
Воспроизводимость результатов экспериментов	Дисперсия воспроизводимости результатов экспериментов $s^2_{\text{восп}}$	0,0015	0,0016	0,0006	0,0015	0,0018	0,0006
	Среднеквадратичная погрешность воспроизводимости $s_{\text{восп}}$	0,0387	0,0400	0,0245	0,0387	0,0424	0,0245
	Уровень значимости α	0,05					
	Число степеней свободы для дисперсии воспроизводимости $f_{\text{восп}}$	32					
Проверка гипотезы об адекватности математической модели	Дисперсия адекватности $s^2_{\text{ад}}$	0,00223	0,00278	0,000957	0,00227	0,00281	0,000996
	Уровень значимости α	0,05					
	Число степеней свободы для дисперсии адекватности $f_{\text{ад}}$	26	27	26	26	27	27
	Экспериментальное значение F -критерия	1,4842	1,7364	1,5946	1,5102	1,5599	1,6599
	Критическое значение F -критерия для адекватности	1,8458	1,8378	1,8458	1,8458	1,8378	1,8378
	Адекватность модели	Адекватная					
Анализ математической модели на информативность	Коэффициент множественной корреляции R	0,9920	0,9905	0,9924	0,9919	0,9904	0,9911
	Уровень значимости α	0,05					
	Число степеней свободы для коэффициентов модели $f_{k'}$	5	4	5	5	4	4
	Число степеней свободы для остаточной суммы квадратов $f_{\text{ост}R}$	26	27	26	26	27	27
	Экспериментальное значение F -критерия для R	322,4	348,6	336,2	317,8	346,3	375,0
	Критическое значение F -критерия для R	2,59	2,73	2,59	2,59	2,73	2,73
	Критерий Бокса и Веца γ	10	10	10	10	10	10
	Информативность модели	Очень высокая					
Число обусловленности $\text{cond}(\mathbf{X}^{\text{T}}\mathbf{X})$		1	1	1,430	1	1	1
Среднее абсолютных величин погрешностей аппроксимации $\lvert\bar{e}_u\rvert$		0,0336	0,0388	0,0232	0,0349	0,0387	0,0235
Доля рассеивания, объясняемая моделью, $Q_{\hat{y}}$		0,9841	0,9810	0,9848	0,9839	0,9809	0,9823

При плохой обусловленности матрицы дисперсий-ковариаций оценки коэффициентов не устойчивы и имеют большие абсолютные величины. Идея регуляризации заключается в увеличении матрицы $\mathbf{X}^{\mathrm{T}}\mathbf{X}$ на неотрицательно определенную матрицу $r\mathbf{E}$, что делает ее более определенной и, как следствие, оценки коэффициентов становятся более устойчивыми [9, с. 118–121]. Новые оценки коэффициентов

$$\mathbf{B}_r = (\mathbf{X}^{\mathrm{T}}\mathbf{X} + r\mathbf{E})^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{Y},$$

где $\mathbf{B}_r$ – матрица-столбец регуляризованных коэффициентов математической модели;

r – параметр регуляризации; $r > 0$;

$\mathbf{X}$ – расширенная матрица эффектов определяемой модели;

$\mathbf{X}^{\mathrm{T}}\mathbf{X}$ – информационная матрица (Фишера);

$\mathbf{E}$ – единичная матрица;

$\mathbf{Y}$ – матрица-столбец полученных результатов опытов;

т – знак транспонирования матрицы $\mathbf{X}$;

–1 – знак обратной матрицы.

Указанное действие приводит к определенным систематическим ошибкам, то есть к смещению коэффициентов модели b_i (что затрудняет интерпретацию коэффициентов в предметной области) и уменьшает их квадратичные ошибки оценок.

Исследования с целью выявления эффективности регуляризации при получении моделей с использованием метода наименьших квадратов показали их хорошую эффективность по сравнению с обычным применением метода наименьших квадратов. Однако при практическом использовании регуляризации (иначе ридж-оценивания) конкретный выбор значения параметра регуляризации r затруднен: аналитическое исследование свойств ридж-оценок труднодоступно (вычислить оптимальные значения r теоретически невозможно [10, с. 69]), так как методология выбора параметра r является стохастической. Определение приближенно оптимального значения $r^*_{опт}$ основано на процедуре, приведенной в [11]. Анализ фактического использования регуляризации в прикладных исследованиях с применением линейного регрессионного анализа показывает, что она используется крайне редко. Общие методы устранения мультиколлинеарности факторов, кроме регуляризации, если невозможно планировать эксперимент, в доступных публикациях отсутствуют.

При получении многофакторных статистических моделей требуется системный подход – план эксперимента, структура модели, элементы структуры модели, форма факторного пространства.

4. Выводы

1. Получение многофакторных статистических моделей проводится в условиях неопределенности свойств моделируемого объекта – системы или процесса.
2. Выбор структуры модели и ее коэффициентов проводится из множества структурных элементов схемы полного факторного эксперимента.
3. Основным условием выбора модели является условие использования расширенной концепции ортогональности и нормирования эффектов при условии возможности планирования эксперимента.
4. Если планировать эксперимент не представляется возможным, то необходимо попытаться получить статистическую модель, используя метод регуляризации, и проверить качество полученной модели.
5. При получении многофакторных статистических моделей следует использовать системный подход.

С разработанными методами решения задач и полученными результатами можно ознакомиться в [12, 13].

СПИСОК ЛИТЕРАТУРЫ

1. Каазик Ю.Я. Математический словарь / Каазик Ю.Я. – М.: ФИЗМАТЛИТ, 2007. – 336 с.
2. Артоболевский И. И. Механика и управление машинами / И.И. Артоболевский // Будущее науки. Международный ежегодник. – 1976. – Вып. 9. – С. 16 – 17.
3. Современные методы оптимизации композиционных материалов / В.А. Вознесенский, В.Н. Выровой, В.Я. Керш [и др.]; под ред. В.А. Вознесенского. – К.: Будівельник, 1983. – 144 с.
4. Тихонов А.Н. Методы решения некорректных задач / А.Н. Тихонов, В.Я. Арсенин. – [3-е изд.]. – М.: Наука, 1986. – 288 с.
5. Тихонов А.Н. Об обратных задачах / А.Н. Тихонов // Некорректные задачи математической физики и анализа. – Новосибирск: Наука, 1984. – С. 161 – 171.
6. Жуковский Е.Л. Статистическая регуляризация решений обратных некорректно поставленных задач обработки и интерпретации результатов эксперимента / Е.Л. Жуковский // Методы математического моделирования, автоматизация обработки наблюдений и их применения: сборник; под ред. А.Н. Тихонова, А.А. Самарского. – М., 1986. – С. 47 – 72.
7. Радченко С.Г. Методология регрессионного анализа / Радченко С.Г. – К.: «Корнійчук», 2011. – 376 с.
8. Орлюк В.А. Технологическое обеспечение собираемости многоэлементных конструкций в условиях роботизированной сборки: автореф. дис. на здобуття наук. ступеня канд. техн. наук: спец. 05.02.08 / В.А. Орлюк. – К.: Киев. политехн. ин-т, 1989. – 16 с.
9. Вучков И. Прикладной линейный регрессионный анализ / Вучков И., Бояджиева Л., Солаков Е.; пер. с болг. и предисл. Ю.П. Адлера. – М.: Финансы и статистика, 1987. – 239 с.
10. Семенов Н.А. Алгоритм расчета гребневых оценок параметров уравнения множественной линейной регрессии / Н.А. Семенов, В.Л. Ракчеева // Автоматика. – 1986. – № 4. – С. 68 – 70.
11. Lennart N. A procedure for determination of a good ridge parameter in linear regression / N. Lennart // Commun. Statist. Simul. and Comput. – 1982. – Vol. 11, N 3. – P. 285 – 309.
12. Лаборатория экспериментально-статистических методов исследований (ЛЭСМИ) [Электронный ресурс]. – Режим доступа: http://www.n-t.org/sp/lesmi.
13. Сайт кафедры «Технология машиностроения» Механико-машиностроительного института Национального технического университета Украины «Киевский политехнический институт» [Электронный ресурс]. – Режим доступа: http://tm-mmi.kpi.ua/index.php/ru/1/publications.

Стаття надійшла до редакції 16.12.2015