

Эффективная концепция регрессионного анализа

С.Г. Радченко

КПІ ім. Ігоря Сікорського, м. Київ, Україна

Received: 24 May 2017 / Accepted: 15 September 2017

Аннотация. Исследована эффективная концепция регрессионного анализа при моделировании сложных систем и процессов. Целью данной работы является использование метода формализованного получения структуры многофакторной регрессионной модели и устойчивого оценивания ее коэффициентов для построения высокоточных статистических моделей характеристик процесса трения трущейся пары поршень-цилиндр в двигателях внутреннего сгорания.

Предполагается, что структура модели исследователю не известна. Аппроксимация исходных данных осуществляется с использованием полиномиальных математических моделей. В качестве теоретического основания корректного статистического моделирования используется понятие полного факторного эксперимента. При невозможности проведения полного факторного эксперимента рекомендуется использовать многофакторные регулярные планы и планы на основе ЛП_r равномерно распределенных последовательностей. Предложена расширенная концепция ортогональности получаемой модели: план эксперимента, структура модели и структурные элементы модели ортогональны друг к другу. Ортогональная структура многофакторной статистической модели позволяет получить статистически независимые оценки коэффициентов моделируемой функции. Такая структура может быть определена однозначно со статистически значимыми коэффициентами. Нормирование ортогональных эффектов позволяет получить максимально устойчивую структуру модели и, следовательно, ее коэффициентов. Решаемая задача будет корректно поставленной.

Рассмотренный метод формализованного получения структуры многофакторной статистической модели и устойчивого оценивания ее коэффициентов использован для построения статистических моделей триботехнических характеристик трения пары поршень-цилиндр двигателя внутреннего сгорания. Получены три модели, характеризующиеся хорошими параметрами качества: адекватностью, информативностью, статистической эффективностью, устойчивостью. В структуре моделей факторы представлены ортогональными нормированными контрастами. Использование моделей позволяет провести анализ механизмов происходящих явлений и определить оптимальные условия работы пары поршень-цилиндр в двигателях внутреннего сгорания. Результаты использования расширенной концепции ортогональности при построении моделей характеристик трения пары поршень-цилиндр подтвердили перспективность применения рассматриваемого подхода, его эффективность и целесообразность при построении регрессионных статистических моделей сложных систем и процессов.

Ключевые слова: регрессионный анализ, математическое моделирование, структура модели, система ортогональных нормированных контрастов, статистическая эффективность модели, устойчивость модели.

1. Введение

Математическое моделирование используется при создании и совершенствовании сложных систем и процессов – наукоемких изделий, высоких технологий, интеллектуальных средств измерений и др. Значительная сложность объектов моделирования, их новизна затрудняют применение теоретико-аналитического подхода. Основным подходом является экспериментально-статистический. В качестве исходной информации широко используются результаты проведенных экспериментов. Модели получают с применением регрессионного анализа и метода наименьших квадратов. Необходимая исходная информация о структуре модели, значимо влияющих факторах часто отсутствует.

Получаемые в экспериментах данные являются результатом суммарного влияния управляемых X , неуправляемых Z и неконтролируемых W факторов. Объект моделирования рассматривается как кибернетическая модель сложной системы с фиксированием множества входов – факторов – и множества выходов – критериев качества – моделируемой системы.

При решении задач регрессионного анализа выбор структуры многофакторной модели часто проводится в виде постулирования структуры полинома определенного порядка без обоснования предложенной структуры. Полученная модель может быть неадекватной и неустойчивой. Обращается внимание на «...принципиальное отличие неформализуемой задачи выбора вида искомой математической модели исследуемого явления и чисто формальной задачи нахождения по экспериментальным данным значений параметров уже выбранной аппроксимирующей функции» [1, с. 277].

Д.ф.-м.н. С.А. Айвазян, д.ф.-м.н. И.С. Енюков и д.ф.-м.н. Л.Д. Мешалкин отмечают, что «на первый план выходит задача правильного определения структуры модели (т. е. выбора общего вида функции $f(X)$), решение которой обеспечивает возможность количественного измерения эффекта воздействия на $Y(X)$ каждой из объясняющих переменных $x^{(1)}, x^{(2)}, \dots, x^{(p)}$ в отдельности. Однако как раз это место (правильный выбор общего вида функции $f(X)$) и является самым слабым во всей технике статистического исследования зависимостей: к сожалению, не существует стандартных приемов и методов, которые образовывали бы строгую теоретическую базу для решения этой важнейшей задачи» [2, с. 22].

В [3, с. 383] отмечается, что «...этап исследования, посвященный выбору общего вида функции регрессии (параметризация модели), бесспорно является ключевым: от того, насколько удачно он будет реализован, решающим образом зависит точность восстановления неизвестной функции регрессии $f(X)$. В то же время приходится признать, что этот этап находится, пожалуй, в самом невыгодном положении: к сожалению, не существует системы стандартных рекомендаций и методов, которые образовали бы строгую теоретическую базу для его наиболее эффективной реализации».

В [3] не приведены рекомендации по выбору плана эксперимента для получения модели. Предлагается использовать априорную информацию о содержательной сущности анализируемой зависимости; предварительный анализ геометрической структуры исходных данных; различные статистические приемы обработки исходных данных, позволяющие сделать наилучший выбор из нескольких сравниваемых вариантов.

Н.П. Тихомиров и Е.Ю. Дорохина отмечают, что «...в практике экономических исследований используется достаточно широкий круг функциональных зависимостей между переменными» [4, с. 18–19]. Приводятся восемь наиболее часто используемых типов зависимостей между переменными y_j и x_i [4, с. 19].

Обращается внимание на возможность ошибки двух видов при обосновании модели: выбрана неправильная форма функционала $f(\alpha, x_i)$ и неверно определен состав ее независимых переменных. Рассматриваются последствия неправильного выбора. Рекомендации по формализованному получению спецификации формы уравнения модели не приводятся [4, с. 83–86].

В книге [5] представлено полное классическое введение в фундаментальные основы регрессионного анализа. Описываются методы подбора и исследования линейных и нелинейных регрессионных моделей различной сложности, а так же рассматриваются практические аспекты их применения, в том числе с использованием специальных компьютерных программ. Помимо стандартного набора тем, составляющих ядро курса регрессионного анализа, в третье издание включены отдельные главы, посвященные мультиколлинеарности, обобщенным линейным моделям, геометрическим свойствам регрессии, робастной регрессии и процедурам тиражирования выборки (бутстрепа). Мало внимания уделено системе ортогональных полиномов Чебышева, методам теории планирования эксперимента; не приведена информация о многофакторных планах экспериментов, планах на основе ЛП_r равномерно распределенных последовательностях. Не рассматриваются свойства полного факторного эксперимента. Не приведена концепция ортогональности в регрессионном анализе как одна из основных идей. Понятие корректности в регрессионном анализе не рассматривается.

В книге [6] рассматриваются способы анализа наблюдений методами математической статистики. Излагаются современные методы анализа распределений вероятностей, оценки связей между случайными величинами, проверки статистических гипотез. Основное внимание посвящено пояснению примеров применения методов современной математической статистики. Регрессионному анализу посвящен один раздел. Круг рассматриваемых вопросов ограничен. По планированию эксперимента изложены начальные элементарные сведения. Проблемы статистической эффективности получаемых моделей, устойчивости структуры и коэффициентов не рассматриваются. Понятие корректности в регрессионном анализе не рассматриваются.

Во многих публикациях по регрессионному анализу не рассмотрены проблемы некорректно поставленных задач, устойчивости получаемых статистических моделей, проверки их на устойчивость и по другим критериям качества модели [7, 8].

В монографии [9] показано, что применение эвристических методов наряду с формализованными расширяет спектр решений в регрессионном анализе. Приведены эвристические алгоритмы получения эвристических планов экспериментов и структур моделей для нестандартных условий проведения экспериментов.

При оптимизация светодиодной системы освещения витаминной космической оранжереи [10, 11] использовалось оптимальное планирование эксперимента и формализованное получение структуры

регрессионной модели, что позволило получить адекватные, устойчивые и информативные модели.

2. Постановка задачи, цель статьи

Целью данной работы является использование метода формализованного получения структуры многофакторной регрессионной модели и устойчивого оценивания ее коэффициентов для построения высокоточных статистических моделей характеристик процесса трения трущейся пары поршня и цилиндра в двигателях внутреннего сгорания.

3. Решение проблемы

Использование математического метода в прикладном исследовании должно быть корректным. Задача считается корректно поставленной, если при заданных условиях решение существует, решение единственное, при сравнительно малых изменениях исходных данных получаемые результаты изменяются сравнительно незначительно. Нарушение хотя бы одного из требований приводит к тому, что задача является некорректно поставленной и ее решение не имеет объективной ценности.

Для успешного математического моделирования в регрессионном анализе необходимо разработать метод получения структуры многофакторной статистической модели не известной исследователю заранее, статистически независимое (ортогональное) оценивание коэффициентов модели, устойчивое оценивание получаемой модели, рекомендации по выбору плана эксперимента. Разработанные методы должны учитывать прикладное использование решений и быть универсальными.

Аппроксимацию исходных данных будем осуществлять с использованием полиномиальных математических моделей. Их применение обосновано теоремами Вейерштрасса, Стоуна, Джексона [12, с. 87–88].

В качестве теоретического основания корректного математического моделирования в регрессионном анализе предлагается использовать понятие полного факторного эксперимента. По теореме Бродского В.З. [13, с. 26–29] все эффекты полного факторного эксперимента ортогональны друг к другу. Полный факторный эксперимент – эксперимент, включающий все возможные комбинации уровней изучаемых факторов при выбранном числе уровней по каждому фактору. Число опытов $N_{\Pi} = s_1^{k_1} \times s_2^{k_2} \times \dots \times s_k^{k_k}$, где s_1, s_2, \dots, s_k – число различных уровней для числа факторов k_1, k_2, \dots, k_k .

Анализ условий решения прикладных реальных задач, свойств полного факторного эксперимента и многофакторных регулярных планов позволили сформулировать основные требования к структуре статистической модели.

1. Структурная группа коэффициентов многофакторного уравнения регрессии не известна исследователю.

2. Структуры моделей выбираются из структуры модели полного факторного эксперимента с ортогональными или близкими к ортогональным структурными эффектами.

3. Выбор структуры модели должен быть формализованным.

4. Возможность формализованного отображения в математической модели произвольной (но конечной) по сложности реальной действительности при условии правильного выбора степени полинома по каждому фактору.

5. Доступность, простота и надежность фактического получения адекватной структуры при решении задач на потоке.

Предложена расширенная концепция ортогональности получаемой модели: план эксперимента, структура модели и структурные элементы модели ортогональны друг к другу. Ортогональная структура многофакторной статистической модели позволяет получить статистически независимые оценки коэффициентов моделируемой функции.

Формализованная структура многофакторной статистической модели задается выражением

$$\prod_{i=1}^k (1 + x_i^{(1)} + x_i^{(2)} + \dots + x_i^{(s_i-1)}) \rightarrow N_{\Pi}$$

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

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

s_i – число различных уровней фактора X_i ;

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

(1), (2), ..., (s-1) – порядок контрастов фактора X_i ;

N_{Π} – число структурных элементов полного факторного эксперимента, равное числу опытов эксперимента.

Число структурных элементов является достаточным для независимого оценивания всех эффектов: главных и взаимодействий. Все эффекты необходимо выразить в виде ортогональных нормированных контрастов. Тогда матрица дисперсий-ковариаций будет диагональной. Модель будет соответствовать

наилучшим возможным критериям D -, A -, E -, G -оптимальности, ортогональности и адекватна результатам экспериментов.

В общем случае полный факторный эксперимент имеет значительное число опытов и при решении реальных прикладных задач не всегда выполним. Необходимо использовать дробный факторный эксперимент в виде многофакторных регулярных планов, которые близки по статистическим свойствам полному факторному эксперименту, и планы на основе ЛП_т равномерно распределенных последовательностей [12, с. 93–111], которые являются квазиполными факторными экспериментами.

Общая методика проведения многофакторного статистического моделирования приведена в [12, с. 209–210; 14].

Коэффициенты многофакторной статистической модели должны быть устойчивы по отношению погрешностей исходных данных, а также погрешностей измерений и вычислений с исходными данными. Устойчивость коэффициентов оценивается критериями Неймана-Голдштейна:

$$P(\mathbf{X}^T \mathbf{X}) = \lambda_{\max}/\lambda_{\min},$$

где \mathbf{X} – расширенная матрица эффектов уравнения регрессии;

$\lambda_{\max}/\lambda_{\min}$ – максимальное и минимальное собственные числа для информационной матрицы Фишера $\mathbf{X}^T \mathbf{X}$;

и числом обусловленности матрицы $\mathbf{X}^T \mathbf{X}$ cond

$$\text{cond}(\mathbf{X}^T \mathbf{X}) = \|\mathbf{X}^T \mathbf{X}\| \times \|(\mathbf{X}^T \mathbf{X})^{-1}\|,$$

где $\|\cdot\|$ – обозначение нормы матрицы.

Наилучшие значения для обоих критериев будут равны 1. Устойчивость хорошая $1 < \text{cond}(\mathbf{X}^T \mathbf{X}) \leq 10$, удовлетворительная $10 < \text{cond}(\mathbf{X}^T \mathbf{X}) \leq 100$, неудовлетворительная $\text{cond}(\mathbf{X}^T \mathbf{X}) > 100$.

Обратим внимание на важность ортогонального представления и нормирования эффектов для устойчивого оценивания коэффициентов регрессионных моделей. Рассмотрим пример. План эксперимента $3^2/9$ представлен в табл. 1.

Таблица 1

План эксперимента $3^2/9$

Номер опыта	X_1	X_2
1	0,1	20
2	0,2	20
3	0,3	20
4	0,1	50
5	0,2	50
6	0,3	50
7	0,1	80
8	0,2	80
9	0,3	80

Структура модели в ортогональных контрастах главных эффектов будет иметь вид

$$\hat{y} = x_1 + z_1 + x_2 + z_2$$

$$x_1^{(1)} = x_1 = 12,247449(X_1 - 0,2); \quad x_2^{(1)} = x_2 = 0,040825(X_2 - 50);$$

$$x_1^{(2)} = z_1 = 1,414214(x_1^2 - 1); \quad x_2^{(2)} = z_2 = 1,414214(x_2^2 - 1).$$

Нормированные ортогональные контрасты представлены в табл. 2.

Таблица 2

Нормированные ортогональные контрасты

Номер опыта	x_1	z_1	x_2	z_2
1	-1,224745	0,707107	-1,224745	0,707107
2	0	-1,414214	-1,224745	0,707107
3	1,224745	0,707107	-1,224745	0,707107
4	-1,224745	0,707107	0	-1,414214
5	0	-1,414214	0	-1,414214
6	1,224745	0,7071067	0	-1,414214
7	-1,224745	0,7071067	1,224745	0,707107
8	0	-1,414214	1,224745	0,707107
9	1,224745	0,707107	1,224745	0,707107

Коэффициенты парной корреляции между всеми эффектами равны нулю:

$$r_{ij}(x_1, z_1) = 0; \quad r_{ij}(x_2, z_2) = 0.$$

Критерии устойчивости

$$P(\mathbf{X}^T \mathbf{X}) = \lambda_{\max} / \lambda_{\min} = 9/9 = 1;$$

$$\text{cond}(\mathbf{X}^T \mathbf{X}) = 1.$$

Натуральные значения ненормированных эффектов представлены в табл. 3

Таблица 3

Натуральные значения ненормированных эффектов

Номер опыта	X_1	X_1^2	X_2	X_2^2
1	0,1	0,01	20	400
2	0,2	0,04	20	400
3	0,3	0,09	20	400
4	0,1	0,01	50	2500
5	0,2	0,04	50	2500
6	0,3	0,09	50	2500
7	0,1	0,01	80	6400
8	0,2	0,04	80	6400
9	0,3	0,09	80	6400

Коэффициенты парной корреляции между эффектами

$$r_{ij}(X_1, X_1^2) = 0,9897; \quad r_{ij}(X_2, X_2^2) = 0,9853.$$

Критерии устойчивости

$$P(\mathbf{X}^T \mathbf{X}) = \lambda_{\max} / \lambda_{\min} = 1,42136 \cdot 10^8 / (4,5789 \cdot 10^{-4}) = 3,1041 \cdot 10^{11};$$

$$\text{cond}(\mathbf{X}^T \mathbf{X}) = 3,791 \cdot 10^{11}.$$

Во втором случае число обусловленности $\text{cond}(\mathbf{X}^T \mathbf{X})$ возросло более чем в 10^{11} раз, хотя исходный план эксперимента был один и тот же.

Вывод. Помимо ортогонального плана эксперимента необходимо обеспечить ортогональное представление эффектов и их нормирование.

Отметим, что при вычислении собственных чисел и собственных векторов по различным алгоритмам и программам для плохо обусловленных матриц могут быть получены результаты, отличающиеся между собой. В руководствах по программному обеспечению встречаются предостережения, что используемые системы MathCad и Maple не позволяют вычислить собственные числа и собственные векторы некоторых, плохо обусловленных, матриц.

4. Пример построения статистических моделей

Исследовалась триботехнология формирования поверхностей трущейся пары поршень-цилиндр двигателя внутреннего сгорания. Изучалось влияние на триботехнические характеристики сопряжения параметров процесса трения: скорости относительного перемещения в контакте $V(F_1)$ м/с, нагрузки в контакте пары трения $L(F_2)$ Н, температуры поверхности трения $T(F_3)$ °С. Для исследования использовался дробный многофакторный регулярный план $3^2 \times 4 // 12$. Уровни варьирования влияющих режимов процесса трения приведены в табл. 4. Триботехнические характеристики пары оценивались по моменту трения в сопряжении $M_{\text{тр}}(y_1)$, коэффициенту трения $d(y_2)$ и интенсивности износа обработанной поверхности $I_h(y_3)$.

Исследование проводилось на закаленном чугуна ($HRC 48 \dots 52$) при трении со смазкой. Окончательная обработка поверхностей проводилась методом антифрикционно-деформационного хонингования. Использовался антифрикционный брусок КМ2/1-МО8-1Ц.

Математические модели оцениваемых триботехнических характеристик постулировались в виде ортогональных многочленов Чебышева. Структура модели полного факторного эксперимента имеет вид

$$(1 + x_1^{(1)} + x_1^{(2)})(1 + x_2^{(1)} + x_2^{(2)})(1 + x_3^{(1)} + x_3^{(2)} + x_3^{(3)}),$$

где $x_1^{(1)}, x_2^{(1)}, x_3^{(1)}$ – линейные ортогональные контрасты факторов X_1, X_2, X_3 ;

$x_1^{(2)}, x_2^{(2)}, x_3^{(2)}$ – квадратичные ортогональные контрасты факторов X_1, X_2, X_3 ;

$x_3^{(3)}$ – кубический ортогональный контраст фактора X_3 .

Из элементов структуры полного факторного эксперимента выбирались элементы структуры дробного факторного эксперимента $3^{2 \times 4 / 12}$.

Рабочая матрица исследования процесса трения, значения уровней факторов и результаты опытов приведены в табл. 4. При проведении эксперимента каждый опыт повторялся дважды.

Таблица 4

Рабочая матрица и результаты опытов

Кодированные теоретические значения уровней варьирования факторов			Факторы			Функции					
			Натуральное обозначение факторов			Натуральное обозначение функций					
			V	L	T	$M_{тр}$		d		I_h	
			Кодированное обозначение факторов и натуральные значения их уровней			Кодированное обозначение функций					
F_1	F_2	F_3	X_1	X_2	X_3	y_1		y_2		y_3	
0	0	0	0,9	115	30	Результаты повторных опытов					
1	1	1	1,5	245	105						
2	2	2	2,9	375	180						
–	–	3	–	–	255	y_{11}	y_{12}	y_{21}	y_{22}	y_{31}	y_{32}
Опыт 1			0,9	115	30	0,434	0,525	0,143	0,173	1,392	0,871
2			0,9	245	180	0,668	0,824	0,105	0,098	4,382	3,291
3			0,9	115	255	0,551	0,512	0,182	0,169	2,617	1,768
4			0,9	245	105	0,520	0,551	0,081	0,087	2,907	1,839
5			1,5	115	30	0,460	0,545	0,152	0,180	2,404	2,096
6			1,5	375	180	0,525	0,564	0,054	0,058	4,011	2,991
7			1,5	115	255	0,460	0,486	0,152	0,160	2,205	1,537
8			1,5	375	105	0,551	0,486	0,057	0,050	3,497	2,339
9			2,9	245	30	0,369	0,336	0,058	0,052	2,493	1,710
10			2,9	375	180	0,980	0,850	0,101	0,088	12,739	9,262
11			2,9	245	255	0,759	0,720	0,119	0,113	30,879	17,541
12			2,9	375	105	0,655	0,915	0,068	0,095	25,415	30,338

Вычисление статистических моделей момента трения \hat{y}_1 , коэффициента трения \hat{y}_2 , интенсивности износа обработанной поверхности \hat{y}_3 и всех критериев качества моделей было проведено с использованием ПС ПРИАМ. Полученные математические модели имеют вид

$$\hat{y}_1 = [58,433 + 29,5662344 x_1^{(1)} - 41,9712245 x_2^{(1)} + 2,74030417 x_2^{(2)} + 5,97775578 x_3^{(1)} - 40,0948942 x_3^{(2)} - 2,56031714 x_3^{(3)} + 5,21387607 x_1^{(1)} x_3^{(1)} + 2,36107074 x_1^{(1)} x_3^{(3)} - 8,85034204 x_1^{(2)} x_2^{(1)} + 3,64117287 x_3^{(1)} x_1^{(2)}] \cdot 10^{-2};$$

$$\hat{y}_2 = [10,887 + 1,06037371 x_1^{(2)} - 3,42976414 x_2^{(1)} + 1,91256702 x_2^{(2)} + 0,888836011 x_3^{(1)} + 0,7507056 x_1^{(1)} x_3^{(1)} + 0,729960576 x_1^{(2)} x_2^{(1)} + 0,694802776 x_1^{(2)} x_3^{(1)}] \cdot 10^{-2};$$

$$\hat{y}_3 = [7,084303 + 6,34934594 x_1^{(1)} + 2,36317033 x_1^{(2)} + 1,17025301 x_2^{(2)} + 2,02333772 x_3^{(1)} - 1,49991131 x_3^{(2)} + 2,51664845 x_3^{(3)} + 2,35767472 x_1^{(1)} x_3^{(1)} + 3,83228738 x_1^{(1)} x_3^{(3)} + 0,968844116 x_1^{(2)} x_3^{(1)} + 1,11401421 x_1^{(2)} x_3^{(3)}] \cdot 10^{-9},$$

где

$$x_1^{(1)} = x_1 = 15(X_1 - 1,7666667);$$

$$x_3^{(1)} = x_3 = 0,0266667 (X_3 - 142,5);$$

$$x_1^{(2)} = 0,0835979 (x_1^2 - 5,5949357 x_1 - 158);$$

$$x_3^{(2)} = 0,25 (x_3^2 - 5);$$

$$x_2^{(1)} = x_2 = 0,0076923 (X_2 - 245);$$

$$x_3^{(3)} = 0,4166667 (x_3^3 - 8,2 x_3).$$

$$x_2^{(2)} = 3(x_2^2 - 0,6666667);$$

В формулах ортогональных контрастов к моделям $\hat{y}_1 \dots \hat{y}_3$ ортогональные контрасты не нормированы. При использовании программного средства ПРИАМ нормировочные коэффициенты ортогональных контрастов вводятся в соответствующие коэффициенты математических моделей [12, с. 56].

Статистический анализ построенных моделей приведен в табл. 5.

Таблица 5

Результаты статистического анализа математических моделей

Параметры статистического анализа параметров		Условные обозначения	Значения для модели		
			\hat{y}_1	\hat{y}_2	\hat{y}_3
Проверка гипотезы о воспроизводимости результатов эксперимента	Дисперсия воспроизводимости	$S_{\text{восп}}^2$	0,00568	0,000127	$2,601 \cdot 10^{-10}$
	Среднеквадратичное отклонение	$S_{\text{восп}}$	0,07536	0,011269	$0,51 \cdot 10^{-9}$
	Число степеней свободы для дисперсии воспроизводимости	$f_{\text{восп}}$	12	12	12
	Экспериментальное значение G -критерия	$G^{\text{эксп}}$	0,496	0,296	0,8009
	Критическое значение G -критерия	$G^{\text{крит}}$	0,541	0,541	0,541
	Уровень значимости	α	0,05		
Значение критерия Стьюдента ($\alpha=0,05; f_{\text{восп}} = 12$)		$t^{\text{крит}}$	2,180	2,180	2,180
Проверка гипотезы об адекватности модели	Экспериментальное значение F -критерия	$F^{\text{эксп}}$	<1	<1	<1
	Критическое значение F -критерия	$F^{\text{крит}}$	1,7472	3,2592	1,7472
	Число степеней свободы для адекватности	$f_{\text{ад}}$	1	4	1
	Уровень значимости	α	0,05		
	Адекватность модели		Адекватна		
Анализ полученной модели на информативность	Коэффициент множественной корреляции	R	0,99522	0,99774	0,99727
	Число степеней свободы для коэффициентов модели	$f_{k'}$	10	7	10
	Число степеней свободы для остаточной суммы квадратов	$f_{\text{ост}R}$	13	16	13
	Экспериментальное значение F -критерия	$F^{\text{эксп}}$	10,390	125,967	18,212
	Критическое значение F -критерия	$F^{\text{крит}}$	2,671	2,6572	2,671
	Уровень значимости	α	0,05		
	Критерий Бокса и Веса	γ	1	3	1
	Информативность модели		удовлетворит.	высокая	удовлетворит.
Число обусловленности		$\text{cond}(\mathbf{X}^T\mathbf{X})$	1,88	1,44	1,92
Среднее абсолютных величин относительных погрешностей аппроксимации		$ \bar{e}_{\text{иотн}} , \%$	2,230	2,930	22,920
Доля рассеяния, объясняемая моделью		Q_j	0,99047	0,99548	0,99454

Рассматриваемые модели адекватны, статистически эффективны, устойчивы, информативны: коэффициент множественной корреляции весьма близок к 1 и значим. Модели объясняют более 99 % рассеивания результатов опытов. Модели зависимости триботехнических характеристик – момента и коэффициента трения и интенсивности износа обработанной поверхности – от условий работы – скорости относительного скольжения и нагрузки в контакте пары и температуры поверхности – отражают физическую сущность процесса трения в интервалах варьирования исследуемых факторов.

С учетом полученных статистических моделей можно провести анализ влияния изменяющихся условий работы на триботехнические характеристики поверхностей трущейся пары поршень-цилиндр и использовать

полученную информацию для определения оптимальных условий ее работы в двигателях внутреннего сгорания.

С разработанными методами моделирования и полученными результатами можно ознакомиться в [15, 16].

5. Выводы

1. Для статистического моделирования с наилучшими критериями качества необходимо обеспечить корректные условия проведения эксперимента

2. Построение многофакторных статистических моделей требует системной организации эксперимента в виде плана эксперимента, структуры модели, элементов структуры модели.

3. Множество результатов опытов должно выражаться в виде полного факторного эксперимента, многофакторных регулярных планов, планов на основе ЛП_т равномерно распределенных последовательностей.

4. Для получения наилучших статистических свойств модели необходимо обеспечить ортогональность структуры модели, структурных элементов модели, использовать ортогональные контрасты и нормирование эффектов модели.

5. Концепция ортогональности структуры статистической модели, использованная при исследовании триботехнических характеристик трущейся пары поршень-цилиндр двигателей внутреннего сгорания, позволила построить адекватные, информативные, устойчивые модели момента трения \hat{y}_1 , коэффициента трения \hat{y}_2 и интенсивности износа обработанной поверхности \hat{y}_3 в зависимости от влияющих факторов: скорости относительного скольжения и нагрузки в контакте пары, температуры поверхности.

6. Результаты использования изложенной концепции регрессионного анализа при моделировании сложных систем и процессов подтвердили ее работоспособность и эффективность.

Ефективна концепція регресивного аналізу

С.Г. Радченко

***Анотація.** Досліджено ефективну концепцію регресійного аналізу при моделюванні складних систем і процесів. Метою даної роботи є використання методу формалізованого отримання структури багатofакторної регресійної моделі і стійкого оцінювання її коефіцієнтів для побудови високоточних статистичних моделей характеристик процесу тертя ?тертьової пари поршень-циліндр в двигунах внутрішнього згорання.*

Передбачається, що структура моделі досліднику не відома. Апроксимація первинних даних здійснюється з використанням поліноміальних математичних моделей. Як теоретична основа коректного статистичного моделювання використовується поняття повного факторного експерименту. При неможливості проведення повного факторного експерименту рекомендовано використовувати багатofакторні регулярні плани і плани на основі ЛП_т рівномірно розподілених послідовностей. Запропоновано розширену концепцію ортогональності одержуваної моделі: план експерименту, структура моделі та структурні елементи моделі ортогональні один до одного. Ортогональна структура багатofакторної статистичної моделі дозволяє отримати статистично незалежні оцінки коефіцієнтів модельованої функції. Така структура може бути визначена однозначно зі статистично значущими коефіцієнтами. Нормування ортогональних ефектів дозволяє отримати максимально стійку структуру моделі і, отже, її коефіцієнтів. Розв'язувана задача буде коректно поставленою.

Розглянутий метод формалізованого отримання структури багатofакторної статистичної моделі і стійкого оцінювання її коефіцієнтів використано для побудови статистичних моделей триботехнічних характеристик тертя пари поршень-циліндр двигуна внутрішнього згорання. Отримано три моделі, що характеризуються хорошими параметрами якості: адекватністю, інформативністю, статистичною ефективністю, стійкістю. У структурі моделей фактори представлені ортогональними нормованими контрастами. Використання моделей дозволяє провести аналіз механізмів явищ, що відбуваються, і визначити оптимальні умови роботи пари поршень-циліндр в двигунах внутрішнього згорання. Результати використання розширеної концепції ортогональності при побудові моделей характеристик тертя пари поршень-циліндр підтвердили перспективність застосування розглянутого підходу, його ефективність і доцільність при побудові регресійних статистичних моделей складних систем і процесів.

***Ключові слова:** регресійний аналіз, математичне моделювання, структура моделі, система ортогональних нормованих контрастів, статистична ефективність моделі, стійкість моделі.*

Effective conception of regression analysis

S. Radchenko

***Abstract.** The paper deals with the efficient conception of regression analysis, when modeling complex systems and processes. The work objective is the use of the method of formalized obtaining of the structure of multifactor regression model and stable estimation*

of its coefficient for construction of highly precise statistical models of the process of friction of a rubbing pair of a piston-cylinder in the internal combustion engine.

It is supposed that the model structure is unknown for the author. Approximation of initial data is performed with the use of polynomial mathematical models. The idea of complete factor experiment is used as a theoretical ground of correct statistical modeling. The use of the complete factor experiment being impossible, it is recommended to use the multifactor regular plans on the basis of J_{III} , of uniformly distributed sequences. An extended conception of orthogonality of the obtained model has been proposed: the experiment design, model structure and structure elements of the model orthogonal to each other. The orthogonal structure of the multifactor statistical model allows obtaining statistically independent estimates of coefficients of the modeled function. Such a structure may be defined unambiguously with statistically significant coefficients. Normalization of orthogonal effects allows obtaining a maximally stable model structure, and, consequently, its coefficients. The solved problem will be well-posed

The considered method of formalized obtaining of the structure of multifactor statistical model and stable estimation of its coefficients is used for constructing statistical models of tribotechnical characteristics of friction parameters of a pair cylinder-piston of the internal combustion engine. Three models have been obtained, which are characterized by good quality parameters: adequacy, informativeness, statistical efficiency, stability. The factors in the model structure are presented by orthogonal normalized contrasts. The models application allows analyzing the mechanisms of the going on phenomena and determining optimal operation conditions of the pair of piston-cylinder internal combustion engine. The results of using the extended conception of orthogonality, when constructing the models of friction parameters of the pair piston-cylinder, have confirmed the promising character of application of the considered approach, its efficiency and expediency in the design of regression statistical model of complex systems and processes.

Keywords: regression analysis, mathematical modeling, model structure, system of orthogonal normalized contrasts, statistical efficiency of the model, model stability.

References

1. Novickiy, P.V., Zograf, I.A. (1991), *Ocenka pogreshnostej rezultatov izmerenij* [Assessment of uncertainties of measurement results], Jenergoatomizdat, Leningrad, Russia.
2. Ajvazjan, S.A., Enyukov, I.S. and Meshalkin, L.D. (1985), *Prikladnaja statistika. Issledovnij zavisimostej* [Applied statistics. Explore dependencies], Finans i statistika, Moskow, Russia.
3. Ajvazjan, S.A. and Mhitarjan, V.S. (2001), *Prikladnaja statistika. Osnovi jekonometriki. Teorija verojatnostej i prikladnaja statistika* [Applied statistics. The basics of econometrics. Vol.2, no 1, 2 nd. Probability theory and applied statistics], YNITI-DANA, Moskow, Russia.
4. Tihomirov, N.P. and Dorohina, E.Yu. (2007), *Jekonometrika* [Econometrics], Publishing "Jekzamen", Moskow, Russia.
5. Drejper, N.R. and Смит, Н. (2007), *Prikladnoj regressionnij analiz* [Applied regression analysis], izdatelskij dom "Wiljams", Moskow, Russia.
6. Kobzar, A.I. (2006), *Prikladnaja matematicheskaja statistika. Dlja inzhenerov i nauchnih rabotnikov* [Applied mathematical statistics. For engineers and scientists], FIZMATLIT, Moskow, Russia.
7. Hinkelmann, K. and Kempthorne, O. (2007), *Design and Analysis of Experiments, Introduction to Experimental Design*. 2 nd edn., Wiley-Interscience, Vol. 1, (Wiley Series in Probability and Statistics).
8. Malov, S.V. (2013), *Regressionnij analiz. Teoreticheskie osnovi i prakticheskie rekomendacii* [Regression analysis. Theoretical bases and practical recommendations], Publishing St. Petersburg State University, St. Petersburg, Russia.
9. Radchenko, S.G. (2015), *Formalizovannie i jevrsticheskie reshenija v regressionnom analize* [Formalized and heuristic solutions in regression analysis], Kornijchuk, Kyiv, Ukraina.
10. Konovalova I., Berkovich Yu.A., Erokhin A., and dr. (2016), Modelling of salad plants growth and physiological status in vitamin space greenhouse during lighting regime optimization, *41th COSPAR Scientific Assembly*, Istanbul, Turkey.
11. Konovalova, I.O., Berkovich, Yu.A., Erohin, A.N., and dr. (2016), *Optimizacija svetodiodnoj sistemi osveshchenija vitaninnoj kosmicheskoy oranzherei* [Optimization of the LED lighting system vitamins space greenhouse] *Aviakosmicheskaja i jekologicheskaja medicina*. Vol. 50, no. 3, pp. 17–22.
12. Radchenko, S.G. (2011), *Metodologija regressionnogo analiza* [Methodology of regression analysis], Kornijchuk, Kyiv, Ukraina.
13. Brodsky, V.Z. (1976), *Vvedenie v faktornoe planirovanie jeksperimenta* [Introduction to the factor design of the experiment], Nauka, Moskow, Russia.
14. Radchenko, S.G. (2015), *Analiz metodov modelirovanija slozhnih sistem* [Analysis of methods of the model of complex systems], *Matematichni mashini i sistemi*, no 4. pp. 123–127.
15. Laboratorija jeksperimental'no-statisticheskikh metodov issledovanij (LESMI) [Laboratory of experimental-statistical methods of research], available at: <http://www.n-t.org/sp/lesmi>
16. Sajt kafedri "Tehnologija mashinostroenija" Mehaniko-mashinostroitel'nogo instituta Nacional'nogo tehničeskogo universiteta Ukraini "Kievskij politehničeskij institute" [Department of Machine Building Technology, Mechanics and Machine Building Institute of the National Technical University of Ukraine "Kyiv Polytechnic Institute"], available at: <http://tm-mmi.kpi.ua/index.php/ru/1/publications>