УДК 519.233.5

Радченко С.Г. д.т.н.
КПИ им. Игоря Сикорского, г. Киев, Украина

СТАТИСТИЧЕСКОЕ МОДЕЛИРОВАНИЕ СЛОЖНЫХ СИСТЕМ

Radchenko S.
Igor Sikorsky Kyiv Polytechnic Institute, Kyiv, Ukraine (lapach@ukr.net)

STATISTICAL MODELING OF COMPLEX SYSTEMS

Изложено статистическое моделирование сложных систем с использованием регрессионного анализа для типичных условий решения реальных прикладных задач, когда структура искомой модели исследователю не известна. Показана необходимость использования планирования эксперимента, расширенной концепции ортогональности, системы ортогональных контрастов. Приведены результаты моделирования цифровых весов, которые подтвердили целесообразность предложенного подхода и использованных методов получения моделей. Критерии качества полученной модели характеризуются как наилучшие. Полученная модель адекватна, высокоинформативна, максимально устойчива, семантична, все коэффициенты ортогональны друг к другу. Использование модели позволяет повысить точность средства измерения по критерию средней абсолютной погрешности аппроксимации до 0,012 % – в 13,3 раза, а по критерию среднеквадратичной погрешности аппроксимации до 4,80 – в 11,2 раза. Результаты использования изложенной концепции регрессионного анализа подтвердили ее работоспособность и эффективность. Рассмотренный метод статистического моделирования может успешно применяться при разработке наукоемких объектов, высоких технологий, интеллектуальных средств измерений, в машиностроении, приборостроении, агробиологии и др.
Ключевые слова: *регрессионный анализ; корректные и некорректные задачи, теория планирования эксперимента; статистические модели; система ортогональных контрастов.*

1. Введение

При создании и совершенствовании сложных систем и процессов широко используются математические модели. В качестве исходных данных применяют результаты экспериментов, статистических испытаний, экспертное оценивание, результаты сложных вычислений. Для определения коэффициентов моделей используют регрессионный анализ и метод наименьших квадратов. Модель получают путем аппроксимации исходных данных, которые являются результатом суммарного влияния управляемых, неуправляемых и неконтролируемых факторов. В случае корреляции влияющих факторов, определяемые коэффициенты смешаны с другими коэффициентами. Необходимо по полученным данным $\mathbf{Y}$ восстановить модель $\mathbf{Y} = \mathbf{XB} + \boldsymbol{\varepsilon}$ в условиях присутствия случайных и систематических ошибок $\boldsymbol{\varepsilon}$ ($\mathbf{X}$ – главные эффекты и взаимодействия факторов, $\mathbf{B}$ – коэффициенты статистической модели).

Большинство обратных задач являются некорректно поставленными. Корректными задачами называются классы математических задач, отвечающих некоторым условиям определенности их решений. Задача называется корректной задачей (или корректно поставленной), если выполнены следующие условия (условия корректности): задача имеет решение, каждым исходным данным соответствует только одно решение – однозначность задачи, решение устойчиво. Задача, для которой не выполняется хотя бы одно из условий, характеризующих корректно поставленную задачу, называется некорректно поставленной, или некорректной. Нужно системно проанализировать условия получения модели и найти решение в условиях исходной неопределенности необходимой информации.

Практика решения реальных прикладных задач показала, что некорректно поставленные задачи встречаются весьма часто и разработка методов их решения является актуальной проблемой.

Специалисты в области прикладной математики обращают внимание на актуальность решения прикладных задач и трудности получения хороших результатов. «Самым важным и самым трудным шагом в работе математика в прикладной области является построение математической модели. Как правило, это плод длительных совместных усилий математика и специалиста в соответствующей области. Часто лишь в результате многочисленных бесед и дискуссий удается дать удовлетворяющее обе стороны математическое описание явления» [1, с. 5].

Научное изучение сложной системы технологии машиностроения связано с изучением связей и закономерностей в производственных процессах изготовления машин [2, с. 7]. В качестве формализованного описания используются уравнения связи [2, с. 70]

DOI: http://dx.doi.org/10.20535/2305-9001.2016.78.85997

$$y = f(x_1, x_2, \ldots, x_n),$$

где y – критерий качества системы;

$x_1, x_2, \ldots, x_n$ – факторы моделируемой системы.

Стохастические зависимости изучаются методами корреляционно-регрессионного анализа [3, с. 607]. Методология получения зависимости не излагается. Проводятся простейшие примеры. Выбор плана эксперимента, структура модели не рассматриваются.

В книге [4] изложена философия, теория и практика планирования научных экспериментов. Отдельные главы посвящены вопросам устойчивого планирования, методу Тагути, латинским квадратам; планам 2^k, 2^{k-p}, 3^k, $2^m \times 3^n$; планам второго порядка, планам Бокса-Бенкина. Поиск структуры статистической модели, устойчивость оценивания коэффициентов не рассматривается.

Работа [5] посвящена рассмотрению классической теории и современных результатов в области регрессионного анализа. Изложены основы регрессионного и дисперсионного анализа в предположении нормальности распределения наблюдений, методики выбора статистических гипотез и подбора оптимальной статистической модели с использованием информационных критериев.

В монографии [6] рассмотрены формализованные и эвристические решения в регрессионном анализе. Приведены эвристические алгоритмы получения эвристических планов экспериментов и структур моделей для нестандартных условий проведения экспериментов.

В [7] описано исследование, проведенное для оптимизации светодиодной системы освещения витаминной космической оранжереи. Поиск оптимума осуществляется с использованием экспериментально-статистического подхода и полученной статистической модели, которая характеризуется хорошими параметрами качества, так как было использовано планирование эксперимента.

2. Постановка задачи, цель статьи

Разработка метода формализованного получения множества структурных элементов в виде главных эффектов и взаимодействий главных эффектов структуры многофакторной статистической полиномиальной модели и устойчивого оценивания ее коэффициентов в условиях отсутствия необходимой исходной информации

3. Решение проблемы

Использование регрессионного анализа предполагает выполнение определенной системы предпосылок [8, с. 43–53]. Будем предполагать, что проводимый эксперимент позволяет определенным образом его организовать: выбрать необходимое число уровней факторов, представить эксперимент в виде определенного сочетания уровней факторов между собой. Принято считать, что такой эксперимент является планируемым и критерии качества плана эксперимента могут соответствовать необходимым.

Важной предпосылкой регрессионного анализа является требование статистической и физической независимости управляемых факторов $X_1, \ldots, X_k$

$$r_{ij}(X_i, X_j) = 0; \quad 1 \le i < j \le k,$$

где $r_{ij}(X_i, X_j)$ – коэффициент парной корреляции факторов X_i, X_j.

Хотя всегда говорится о независимости факторов, необходимо и правильно обеспечить независимость всех эффектов – главных эффектов факторов и взаимодействий факторов.

Фактическое обеспечение указанной предпосылки является одной из основных проблем многофакторного регрессионного анализа. Если предпосылка не выполняется, то задача является некорректно поставленной и требует специальных методов ее решения.

Мультиколлинеарность факторов приводит к следующим недостаткам.

1) Ошибки определяемых коэффициентов возрастают и могут достичь значительных величин.

2) Определямые коэффициенты смешаны друг с другом, их величины и в некоторых случаях знаки не соответствуют истинным значениям.

3) Использование статистических критериев (t – Стьюдента, F – Фишера) теоретически не обосновано.

Полученные модели в условиях мультиколлинеарности факторов являются формальными и могут не нести полезной информации о предметной области, которую они описывают.

Будем предполагать, что в качестве плана эксперимента используется полный факторный эксперимент. Формализованная структура многофакторной статистической модели задается выражением

$$\prod_{i=1}^{k}(1 + x_i^{(1)} + x_i^{(2)} + \ldots + x_i^{(s_i-1)}) \to N_{\breve{I}}$$

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

$x_i^{(1)},\ x_i^{(2)}, \ldots, x_i^{(s_i-1)}$ – 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;

$N_{\text{Ï}}$ – число структурных элементов полного факторного эксперимента, равное числу опытов эксперимента.

Предполагается, что порядок максимального значения ортогонального контраста s_i–1 достаточный для адекватного описания влияния фактора X_i по всей области факторного пространства. Значение s_i назначается исследователем, исходя из логически-профессионального анализа предметной области.

Для полного факторного эксперимента число структурных эффектов (элементов) модели равно числу опытов плана эксперимента N_{Π}, и все эффекты ортогональны друг к другу по теореме Бродского В.З. [9, с. 26–29]. Получаемая статистическая модель будет адекватна результатам эксперимента, так как множество структурных элементов необходимо и достаточно для описания результатов опытов.

Предполагается, что число повторных опытов в каждой строке матрицы результатов одинаково для всех строк: $n_u = n$. Для выполнения условий ортогональности эффектов уровни варьирования факторов будем представлять в виде контрастов факторов. Контрастом фактора X_i будем считать любое множество коэффициентов, удовлетворяющее условиям:

$$c_i^{(p)} = f_i^{(p)}(X_i),$$

$$\sum_{u=1}^{N} c_{iu}^{(p)} = 0,$$

$$\sum_{u=1}^{N} \left[c_{iu}^{(p)} \right]^2 \neq 0,$$

где $c_i^{(p)}$ – контраст p-го порядка для X_i фактора, $1 \le p \le s_i - 1$;

$f_i^{(p)}$ – функция p-го порядка от X_i фактора.

Контрасты факторов X_i и X_j – ортогональные контрасты, если

$$\sum_{u=1}^{N} c_{iu}^{(p)} c_{ju}^{(p')} = 0; \quad 1 \le i < j \le k; \quad 1 \le p' \le s_j - 1.$$

При $n_u \neq n$ (число опытов в каждой строке матрицы результатов различно) выражение для контрастов записывают следующим образом:

$$\sum_{u=1}^{N} n_u c_{iu}^{(p)} = 0; \qquad \sum_{u=1}^{N} n_u c_{iu}^{(p)} c_{ju}^{(p')} = 0.$$

Уравнения для нахождения компонентов любых контрастов записывают, исходя из определения контраста фактора X_i

$$\sum_{u=1}^{N} x_{iu}^{(p)} = \sum_{u=1}^{N} f_i^{(p)}(X_{iu}) = 0$$

и условия ортогональности любых контрастов фактора X_i

$$\sum_{u=1}^{N} x_{iu}^{(p)} x_{iu}^{(p')} = \sum_{u=1}^{N} f_i^{(p)}(X_{iu}) f_i^{(p')}(X_{iu}) = 0,$$

где $x_{iu}^{(p)}$ – значение p-го ортогонального контраста i-го фактора для u-той строки матрицы планирования, $1 \le u \le N, \quad 1 \le p \le s_i - 1$;

$x_{iu}^{(p')}$ – значение p'-го ортогонального контраста i-го фактора для u-той строки матрицы планирования, $1 \le p < p' \le s_i - 1$.

Линейные и квадратичные контрасты как функции от уровней варьирования факторов:

$$x_{iu}^{(1)} = x_{iu} = a_{11i}(X_{iu} + a_{10i}),$$

$$x_{iu}^{(2)} = z_{iu} = a_{22i}(x_{iu}^2 + a_{21i} x_{iu} + a_{20i}),$$

где a_{11i}, a_{22i} – коэффициенты, выбираемые после предварительного определения значений контрастов x_{iu}, z_{iu} таким образом, чтобы контрасты имели нормированные значения;

a_{10i}, a_{21i}, a_{20i} – коэффициенты, определяемые из условия ортогональности различных контрастов относительно друг друга для фактора X_i.

Коэффициенты $a_{(.)}$ на 2 уровнях

$$a_{10i} = -\frac{\sum_{u=1}^{N} X_{iu}}{N},$$

на трех уровнях

$$a_{20i} = -\frac{\sum_{u=1}^{N} x_{iu}^2}{N}, \qquad a_{21i} = -\frac{\sum_{u=1}^{N} x_{iu}^3}{\sum_{u=1}^{N} x_{iu}^2}.$$

Все эффекты в полном факторном эксперименте ортогональны друг к другу. Матрица дисперсий-ковариаций будет диагональной, т. е. все внедиагональные элементы равны нулю. Каждый коэффициент вычисляется независимо от других коэффициентов. При нормировании эффектов коэффициенты модели будут максимально устойчивы, $\mathrm{cond}(\mathbf{X}^{\delta}\mathbf{X}) = 1$.

В общем случае использование изложенной концепции затруднено, так как число опытов в полном факторном эксперименте, как правило, весьма велико и невыполнимо. Тогда необходимо использовать дробный факторный эксперимент. В качестве дробного факторного эксперимента нужно использовать многофакторные регулярные планы, которые являются квазиполными факторными экспериментами, и планы на основе ЛП$_\tau$ равномерно распределенных последовательностей [8, с. 93–111]. При вычислении модели необходимо использовать алгоритм RASTA3 и программное средство «Планирование, регрессия и анализ моделей» (ПС ПРИАМ) [8, с. 81–85].

Рассмотренный метод статистического моделирования может успешно применяться при разработке наукоемких объектов, высоких технологий, интеллектуальных средств измерений, в машиностроении, приборостроении, агробиологии и др.

4. Пример повышения точности измерения цифровых весов

Рассмотрим возможности предложенного подхода на примере повышения точности цифровых весов с диапазоном взвешивания 0—100 кгс. Датчик весов емкостного типа с автономным питанием от переносного источника напряжения. Весы предназначены для эксплуатации в диапазоне температуры окружающей среды (воздуха) 0...60 °С. Напряжение от автономного источника напряжения в процессе эксплуатации весов может изменяться в диапазоне 12,3...11,7 В при расчетном (номинальном) значении 12 В.

Предварительное исследование цифровых весов показало, что изменения температуры окружающей среды и питаемого напряжения в вышеприведенных диапазонах сравнительно мало влияют на показания емкостного датчика и, следовательно, на результаты взвешивания. Однако стабилизировать эти внешние и внутренние условия с необходимой точностью и поддерживать их в процессе функционирования весов не представлялось возможным ввиду того, что весы должны эксплуатироваться не в стационарных (лабораторных) условиях, а на борту перемещающегося объекта.

Исследование точности весов без учета влияния изменений температуры и питаемого напряжения показали, что средняя абсолютная погрешность аппроксимации составляет 0,16 %, а среднеквадратичная погрешность остатка (в единицах измерения выходной величины взвешивания) равна 53,92.

Для получения многофакторной математической модели были приняты следующие обозначения факторов.

X_1 — гистерезис. Уровни: 0 (нагрузка); 1 (разгрузка). Фактор качественный.

X_2 — температура окружающей среды. Уровни: 0; 22; 60°С.

X_3 — напряжение питания. Уровни:11,7; 12,0; 12,3 В.

X_4 — измеряемый вес. Уровни: 0; 20; 40; 60; 80; 100 кгс.

Учитывая принятые уровни варьирования факторов и сравнительно не трудоемкий объем испытаний было решено провести полный факторный эксперимент, т. е. $2{\times}3^2{\times}6//108$. Предварительный анализ исходных данных показал, что они со значительной вероятностью содержат грубые ошибки. Эти опыты были повторены и их результаты были исправлены.

Натуральные значения уровней варьирования факторов были преобразованы в ортогональные контрасты, иначе в систему ортогональных полиномов Чебышева.

С использованием системы ортогональных контрастов структура полного факторного эксперимента будет иметь следующий вид.

$$(1 + x_1^{(1)})(1 + x_2^{(1)} + x_2^{(2)})(1 + x_3^{(1)} + x_3^{(2)})(1 + x_4^{(1)} + x_4^{(2)} + x_4^{(3)} + x_4^{(4)} + x_4^{(5)}) \to N_{108},$$

где $x_1^{(1)}, x_2^{(1)}, x_3^{(1)}, x_4^{(1)}, x_2^{(2)}, x_3^{(2)}, x_4^{(2)}, x_4^{(3)}, x_4^{(4)}, x_4^{(5)}$ — соответственно линейные, квадратичные, кубичный, четвертой и пятой степени контрасты факторов X_1, ..., X_4;

N_{108} — число структурных элементов для схемы полного факторного эксперимента.

Все эффекты (главные и взаимодействия) были нормированы

$$\sum_{u=1}^{108}\left(x_{iu}^{(p)}\right)^2 = 108,$$

где $x_{iu}^{(p)}$ — значение p-го ортогонального контраста i-го фактора для u-й строки матрицы планирования, $1 \le u \le 108$, $1 \le p \le s_i - 1$; $1 \le i \le 4$.

Предварительный расчет математической модели показал, что в качестве оценки дисперсии воспроизводимости может быть выбрана (приближенно) величина 20,1.

Число степеней свободы (условно) принято $f_2 = 108$.

Дисперсия была использована для определения стандартной ошибки коэффициентов уравнения регрессии.

Вычисление математической модели и всех ее критериев качества было проведено с использованием ПС ПРИАМ. Полученная математическая модель имеет вид

$$\hat{y} = 28968,9 - 3715,13x_4^{(1)} + 45,2083x_3^{(1)} - 37,5229x_2^{(2)} + 23,1658x_2^{(1)} - 19,0708x_4^{(2)} - 19,6574x_3^{(2)} -$$
$$-9,0094x_2^{(1)}x_3^{(2)} - 9,27434x_2^{(2)}x_4^{(1)} + 1,43465x_1^{(1)}x_2^{(1)} + 1,65431x_2^{(2)}x_3^{(1)}$$

где

$x_1^{(1)} = 2(X_1 - 0,5)$;

$x_2^{(1)} = x_2 = 0,0306122(X_2 - 27,3333)$;

$x_2^{(2)} = 1,96006(x_2^2 - 0,237337\,x_2 - 0,575594)$;

$x_3^{(1)} = x_3 = 3.33333\,(X_3 - 12)$;

$x_3^{(2)} = 1,5(x_3^2 - 0,666667)$;

$x_4^{(1)} = x_4 = 0,02(X_4 - 50)$;

$x_4^{(2)} = 1,875(x_4^2 - 0,466667)$.

В табл. 1 приведены критерии качества полученной многофакторной статистической модели. Модель адекватна. Доля рассеивания, объясняемая моделью, весьма высока, так как модель высокоточная, изменчивость функции отклика велика, а ее случайная изменчивость сравнительно мала. Коэффициент множественной корреляции R весьма близок к 1 и устойчив, так как. будучи скорректированным с учетом степеней свободы, практически не меняется. Статистическая значимость R весьма велика, т. е. модель очень информативна. Высокая информативность модели подтверждается также значением критерия Бокса и Веца.

Таблица 1

Критерии качества полученной статистической модели

Параметры статистического анализа	Значение параметра
АНАЛИЗ АДЕКВАТНОСТИ МОДЕЛИ	
Остаточная дисперсия	21,1084
Дисперсия воспроизводимости (задана пользователем)	20,1
Экспериментальное значение F-критерия	1,05017
Уровень значимости F-критерия для адекватности для степеней свободы $f_1 = 97$ $f_2 = 108$	0,05
Критическое значение F-критерия для адекватности	1,3844
Критическое значение F-критерия (при отсутствии повторных опытов)	1,02681
Стандартная ошибка оценки (скорректированная с учетом степеней свободы)	4,59439 4,80072
Модель АДЕКВАТНА	
АНАЛИЗ ИНФОРМАТИВНОСТИ МОДЕЛИ	
Доля рассеивания, объясняемая моделью	0,999997
Введено регрессоров (эффектов)	11
Коэффициент множественной корреляции R (скорректированный с учетом степеней свободы)	0,999999 0,999998
Экспериментальное значение F критерия для R	$3,29697 \cdot 10^6$
Уровень значимости F-критерия для информативности для степеней свободы $f_1 = 10$; $f_2 = 97$	0,01
Критическое значение F-критерия для информативности	2,50915
Модель ИНФОРМАТИВНА	
Критерий Бокса и Веца для информативности	>49
Информативность модели ОЧЕНЬ ВЫСОКАЯ	

Коэффициенты модели максимально устойчивы: число обусловленности $\mathrm{cond} = 1$. Полученная модель семантична в информационном смысле, так как все ее коэффициенты ортонормированны: они статистически независимы и могут сравниваться по абсолютной величине друг с другом. Знак коэффициента показывает характер влияния, а его абсолютная величина — силу влияния. Полученная модель наиболее удобна для интерпретации в предметной области.

Учитывая семантичные свойства полученной математической модели и доли участия каждого из эффектов модели в общей доле рассеивания, объясняемой моделью, можно провести содержательный информационный анализ формирования результата измерения исследуемых цифровых весов.

Превалирующая доля участия в результатах моделирования, равная 0,999557, создается линейным главным эффектом $x_4^{(1)}$ (с коэффициентом $b_1 = -3715{,}13$), т. е. измеряемым весом (табл. 2). Нелинейность $x_4^{(2)}$ (с коэффициентом $b_5 = -19{,}07$) сравнительно мала ($3{,}16\cdot10^{-5}$) и ее учет в модели повышает точность измерения. Линейный эффект $x_4^{(1)}$ сравнительно слабо ($3{,}19\cdot10^{-6}$) взаимодействует с квадратичным эффектом $x_2^{(2)}$ температуры окружающей среды: взаимодействие $x_2^{(2)}\,x_4^{(1)}$ ($b_8 = -9{,}27$). Следовательно, математическая модель только от фактора измеряемый вес X_4 должна включать и эффект влияния температуры окружающей среды

$$\hat{y}_1 = 28968{,}90 - 3715{,}13\,x_4^{(1)} - 19{,}07\,x_4^{(2)} - 9{,}27\,x_2^{(2)}\,x_4^{(1)},$$

фактор, которого X_2 является неуправляемым.

Напряжение питания изменяет результаты взвешивания в виде линейного эффекта $x_3^{(1)}$ ($b_2 = 45{,}21$) и квадратичного эффекта $x_3^{(2)}$ ($b_6 = -19{,}66$). Их суммарная доля участия составляет $2{,}41\cdot10^{-4}$.

Температура окружающей среды влияет в виде квадратичного $x_2^{(2)}$ (b_3=–37,52) и линейного $x_2^{(1)}$ (b_4=23,17) эффектов с суммарной долей участия $1{,}60\cdot10^{-4}$.

Температура окружающей среды и напряжение питания образуют парное взаимодействие $x_2^{(1)}\,x_3^{(2)}$ (b_7=–9,01) с долей участия $3{,}63\cdot10^{-6}$.

Доказательность статистической значимости двух последних эффектов $x_1^{(1)}\,x_2^{(1)}$ и $x_2^{(2)}\,x_3^{(1)}$ не может быть обоснована, так как они существенно меньше эффектов $x_2^{(1)}\,x_3^{(2)}$ и $x_2^{(2)}\,x_4^{(1)}$, а обоснованное значение дисперсии воспроизводимости по результатам повторных опытов в представленных исходных данных, к сожалению, отсутствовало.

В табл. 2 приведены статистические характеристики коэффициентов регрессии. Отметим, что значения

Таблица 2

Статистические характеристики коэффициентов регрессии

Наименование главного эффекта или взаимодействия главных эффектов	Коэффициент регрессии	Стандартная ошибка коэффициента регрессии	Вычисленное значение t-критерия	Доля участия в объяснении разброса моделируемой величины
$x_4^{(1)}$	$b_1 = -3715{,}13$	0,431406	5882,9	0,999557
$x_3^{(1)}$	$b_2 = 45{,}2083$	0,431406	85,5631	0,000211445
$x_2^{(2)}$	$b_3 = -37{,}5229$	0,431406	62,2275	0,000111838
$x_2^{(1)}$	$b_4 = 23{,}1658$	0,431406	40,7398	$4{,}79362\cdot10^{-5}$
$x_4^{(2)}$	$b_5 = -19{,}0708$	0,431406	33,0808	$3{,}16065\cdot10^{-5}$
$x_3^{(2)}$	$b_6 = -19{,}6574$	0,431406	32,22	$2{,}9983\cdot10^{-5}$
$x_2^{(1)}\,x_3^{(2)}$	$b_7 = -9{,}0094$	0,431406	11,2035	$3{,}62519\cdot10^{-6}$
$x_2^{(2)}\,x_4^{(1)}$	$b_8 = -9{,}27434$	0,431406	10,5069	$3{,}18838\cdot10^{-6}$
$x_1^{(1)}\,x_2^{(1)}$	$b_9 = 1{,}43465$	0,431406	2,523	$1{,}83848\cdot10^{-7}$
$x_2^{(2)}\,x_3^{(1)}$	$b_{10} = 1{,}65431$	0,431406	2,24004	$1{,}44923\cdot10^{-7}$
$b_0 = 28968{,}9$ Уровень значимости для t-критерия 0,05 для степеней свободы $f_1 = 108$ Критическое значение t-критерия 1,98217				

коэффициентов регрессии разделены на нормировочные коэффициенты ортогональных контрастов, которые не включены в приведенные формулы ортогональных контрастов Этим и объясняется то, что при делении значений коэффициентов регрессии на их стандартную ошибку полученные значения t-критерия отличаются от приведенных правильно вычисленных значений этого критерия в табл. 2.

Были проанализированы графики остатков $e_u = \bar{y}_u - \hat{y}_u$. Общий график остатков сравнительно близок к нормальному закону распределения. Временной график остатков указывает на случайный характер изменения остатков от времени (последовательности) проведения опытов. Дальнейшее повышение точности модели не возможно. Анализ зависимости остатков от $\hat{y}$ (расчетного значения) показал приближенно равномерный разброс в зависимости от измеряемого веса.

Учет в математической модели разнообразных систематических погрешностей, нелинейностей, взаимодействий неуправляемых факторов позволил повысить точность средства измерения по критерию средней абсолютной погрешности аппроксимации до 0,012 % – в 13,3 раза, а по критерию среднеквадратичной погрешности аппроксимации до 4,80 (табл. 1) – в 11,2 раза.

Для фактической реализации полученной модели $\hat{y}$ необходимо измерить и использовать информацию о температуре окружающей среды и напряжении питания с помощью датчиков и провести расчет результата с применением микропроцессора.

Итоговый вывод о целесообразности использования изложенного подхода зависит от экономической эффективности следующих сравниваемых вариантов.

1. Высокоточного средства измерения и, следовательно, более дорогого, используемого в нормированных (стандартных) условиях, которые необходимо создать и поддерживать.
2. Средства измерения менее высокой точности, используемого в не нормированных (не стандартных) условиях с применением полученной математической модели.

С разработанными методами решения задач и полученными результатами можно ознакомиться в [10, 11].

5. Выводы

1. Неизвестную структуру многофакторной статистической модели необходимо выбирать из элементов структуры полного факторного эксперимента, что обеспечивает ортогональность всех эффектов друг к другу, статистическую эффективность определяемых коэффициентов модели и их устойчивость.

2. В качестве плана эксперимента желательно использовать полный факторный эксперимент. Если полный факторный эксперимент невозможен, необходимо использовать многофакторный регулярный план или план на основе $ЛП_\tau$ равномерно распределенных последовательностей. План эксперимента не должен быть близким к насыщенному.

3. Натуральные значения уровней варьирования факторов должны быть преобразованы в систему нормированных ортогональных контрастов для обеспечения их наилучших статистических свойств.

4. Необходимо использовать концепцию ортогональности: план эксперимента, структура статистической модели, элементы модели должны быть ортогональны или достаточно близки к ортогональным. Это позволяет создать корректные условия для получения статистической модели.

5. Использование концепции ортогональности позволяет получить модель с критериями качества: адекватности, статистической эффективности, устойчивости, семантичности информационной.

6. Для формирования структуры модели и проведения необходимых преобразований и вычислений следует использовать алгоритм RASTA3 и программное средство «Планирование, регрессия и анализ моделей» (ПС ПРИАМ).

7. Использование изложенной концепции регрессионного анализа при моделировании сложных систем и процессов подтвердило ее работоспособность и эффективность.

8. Успешно реализованный системный подход в математическом моделировании средства измерения позволил учесть влияние факторов внешней – температура окружающей среды – и внутренней среды – напряжение питания. Эффективность извлечения полезной информации из исходных данных составила 100 %.

9. В полученной многофакторной математической модели, структура которой априори исследователю не была известна, в удобной для интерпретации в предметной области форме раскрыты нелинейность средства измерения и системное влияние факторов (эмергентность) внешней и внутренней среды. В реальных условиях эксплуатации стабилизация этих факторов с необходимой точностью не представляется возможной.

10. Учет математической модели систематических погрешностей позволил повысить точность измерений по критерию средней абсолютной погрешности в 13,3 раза и по критерию среднеквадратичной погрешности в 11,2 раза.

Анотація. *Викладено ефективну концепцію регресійного аналізу для типових умов рішення реальних прикладних задач, коли структура шуканої моделі досліднику не відома. Показано необхідність використання планування експерименту, розширеної концепції ортогональності, системи ортогональних контрастів. Критерії якості отриманої моделі характеризуються як найкращі. Отримана модель адекватна, високоінформативна, максимально стійка, семантична, всі коефіцієнти ортогональні один до одного. Використання моделі дозволяє підвищити точність засобу виміру за*

критерієм середньої абсолютної похибки апроксимації до 0,012 % – в 13,3 рази, а за критерієм середньоквадратичної похибки апроксимації до 4,80 – в 11,2 рази. Результати використання викладеної концепції регресійного аналізу підтвердили її працездатність і ефективність. Розглянутий метод статистичного моделювання може успішно застосовуватися при розробці наукоємних об'єктів, високих технологій, інтелектуальних засобів вимірювань,в машинобудуванні, приладобудуванні, агробіології та ін.
Ключові слова: регресійний аналіз; коректні і некоректні задачі; теорія плануванні експерименту; статистичні моделі; система ортогональних контрастів.

***Abstract**. The author expounds statistical modeling of complex systems with the use of regression analysis for typical conditions of solution of real applied problems, when the model structure is unknown for the researcher. A necessity of using the experiment design, extended conception of orthogonality, and a system of orthogonal contrasts is shown. The results of modeling the digital balance are presented which have confirmed the expediency of the offered approach and the methods used for obtaining the models. The quality criteria of the obtained model are characterized as the best ones. The model obtained is adequate, highly informative, maximum stable, semantic, all the coefficients are orthogonal to each other. The model use allows increasing 13.3 times the accuracy of the measuring device by the criterion of the average absolute error to 0.012%, and 11.2 times by the criterion of root-mean-square approximation error to 4.80%. The results of the use of the above stated conception of regression analysis have confirmed its efficiency. The method of statistical modeling can be successfully applied in the development of high-tech facilities, high technology, intelligent measuring instruments, machine building, instrument engineering, agrobiology, etc.*
Keywords: regression analysis; correct and incorrect problems; the theory of experiment planning; statistical models; system of orthogonal contrasts.

References

1. Zhidkov, N.P. (1980), *Predislovie. Chto takoe matematika* [Preface. What is mathematics], Znanie, Moskow, Russia.
2. Kolesov, I.M. (2001), *Osnovy tehnologii mashinostroenija* [Machine building technology principles], Visshaja shkola, Moskow, Russia.
3. Suslov, A.G. and Dalsky, A.M. (2002), *Nauchnye osnovy mashinostroenija* [Scientifi grounds of machine building], Mashinostroenie, Moskow, Russia.
4. Hinkelmann, K. and Kempthorne, O. (2007), *Design and Analysis of Experiments, Introduction to Experimental Design.* 2nd, Wiley-Interscience, Vol. 1, (Wiley Series in Probability and Statistics).
5. 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.
6. Radchenko, S.G. (2015), *Formalizovannie i jevristicheskie reshenija v regressionnom analize* [Formalized and heuristic solutions in regression analysis], Kornijchuk, Kyiv, Ukraina.
7. Konovalova, I.O., Berkovich, Yu.A., Erohin, A.N., ets. (2016), "Optimization of the LED lighting system vitamins space greenhouse", *Aviakosmicheskaja i jekologicheskaja medicina,* Vol. 50, no. 3, pp. 17–22.
8. Radchenko, S.G. (2011), *Metodologija regressionnogo analiza* [Methodology of regression analysis], Kornijchuk, Kyiv, Ukraina.
9. Brodsky, V.Z. (1976), *Vvedenie v faktornoe planirovanie jeksperimenta* [Introduction to the factor design of the experiment], Nauka, Moskow, Russia.
10. Laboratorija jeksperimental'no-statisticheskih metodov issledovanij (LESMI) [Laboratory of experimental-statistical methods of research], available at: http://www.n-t.org/sp/lesmi
11. Sajt kafedri «Tehnologija mashinostroenija» Mehaniko-mashinostroitel'nogo instituta Nacional'nogo tehnicheskogo universiteta Ukraini «Kievskij politehnicheskij institut» [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

Подана до редакції 12.10.2016