

ИСПОЛЬЗОВАНИЕ ЭВРИСТИКИ В ПЛАНИРОВАНИИ ЭКСПЕРИМЕНТА И РЕГРЕССИОННОМ АНАЛИЗЕ

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

***Анотація.** Показана необхідність використання евристичних рішень у статистичних задачах. Наведено розроблені евристики у плануванні експерименту, регресійному аналізі і отримані результати.*

***Ключові слова:** евристика, планування експерименту, регресійний аналіз.*

***Аннотация.** Показана необходимость использования эвристических решений в статистических задачах. Приведены разработанные эвристики в планировании эксперимента, регрессионном анализе и получены результаты.*

***Ключевые слова:** эвристика, планирование эксперимента, регрессионный анализ.*

***Abstract.** A necessity of using heuristic decisions in stochastic problems has been shown. The developed heuristics in experiment design and regression analysis are given; obtained results are presented.*

***Keywords:** heuristics, experiment design, regression analysis.*

1. Введение. Постановка задачи

Моделирование сложных систем – технических, технологических, измерительных – в большинстве случаев проводится с использованием экспериментально-статистического подхода, не отрывая изучаемую систему от определенных рамок времени и места, то есть применяется идиографическое описание. В качестве математических методов используется планирование эксперимента и регрессионный анализ.

Необходимая для описания информация – закон распределения случайных погрешностей, статистически значимо влияющие факторы, структура математической модели – часто отсутствует. Получение моделей проводится в условиях неопределенности. С возрастанием сложности систем, их новизны успешное решение задачи усложняется.

Размерно-геометрические и функциональные параметры систем могут быть коррелированы между собой. Форма факторного пространства может отличаться от многомерного прямоугольного параллелепипеда, сферы, симплекса. В этих случаях приходится применять нестандартные подходы.

Использование традиционных методов планирования эксперимента и регрессионного анализа может быть неэффективным, так как принятые предпосылки могут не выполняться. Приходится разрабатывать новые методы или модифицировать известные. Однако использовать для этого формализованные (математические) решения не всегда возможно ввиду их отсутствия, а разработка новых затруднена.

Цель статьи

В работах автора [1, 2] были разработаны новые методы планирования эксперимента и регрессионного анализа с элементами эвристики, обеспечивающие получение устойчивых решений.

Целью статьи является краткое изложение сути этих методов и представление результатов вычислительного эксперимента.

2. Эвристические решения в планировании эксперимента

Эвристика (с греч. отыскиваю, открываю) – специальные методы решения задач. Под эвристикой понимают организацию процесса продуктивного творческого мышления путем использования интуиции, опыта решения определенного класса задач [3, с. 4–6]. Специалисты, изучающие эвристические методы решения задач, считают, что при этом используется интуитивное мышление. В нем отсутствует четкая структура. Ответ получается без осознания процесса его получения. При аналитическом мышлении отдельные его этапы могут быть сформулированы, приведены в виде алгоритма и формализованы.

Практика решения реальных прикладных задач по моделированию сложных систем показала, что для некоторых сочетаний уровней факторов экономные планы экспериментов в доступных каталогах отсутствуют. При сильном ограничении на число проводимых опытов разработан алгоритм RASTA1 генерирования плана эксперимента с минимально возможной коррелированностью факторов между собой [4, с. 113–115]. За основу плана необходимо взять регулярный план с минимальным числом опытов N , в котором факторы X_{k-1} и $X_{\text{бл}}$ будут изменяться на числе уровней s_{k-1} и $s_{\text{бл}}$. Эвристическое решение состоит в том, что в каждом из блоков вычеркивается опыт с различным значением уровней фактора X_{k-1} . В полученном плане эксперимента с числом опытов $N - s_{\text{бл}}$ коррелированность факторов будет минимально возможной. Необходимо рассчитать коррелированность факторов в полученном плане и сделать вывод о его возможном использовании.

В случае отсутствия необходимого плана эксперимента для определенного числа уровней фактора разработан алгоритм RASTA2 [4, с. 114–115]. В качестве базового плана используется многофакторный регулярный план эксперимента с блоковым фактором $X_{\text{бл}}$ на $s_{\text{бл}}$ уровнях. Эвристика заключается в том, что для фактора $X_{\Phi 1}$ выбирается фактор $X_{\Phi 2}$ с числом уровней $s_{\Phi 2}$, близким к числу уровней $s_{\Phi 1}$ фактора $X_{\Phi 1}$ ($s_{\Phi 2} < s_{\Phi 1}$), для которого построение многофакторного регулярного плана возможно. По фактору $X_{\text{бл}}$ план эксперимента разбивается на $s_{\text{бл}}$ ортогональных блоков. Для фактора $X_{\Phi 2}$ в каждом из блоков заменяются различные не одноименные значения уровней на большее значение уровня, отсутствующего у фактора $X_{\Phi 2}$ и имеющегося у фактора $X_{\Phi 1}$. Вычисляются коэффициенты парной корреляции r_{ij} для всех факторов X_i, X_j полученного плана эксперимента ($1 \leq i < j \leq k; k$ – число факторов). Анализируется корреляционная матрица, и делается вывод о возможном использовании плана.

Приведенные алгоритмы были использованы при получении планов экспериментов $3^3 \times 4^2 \times 9^1 \times 4^1 // 32$ (из плана $3^3 \times 4^2 \times 8^1 \times 4^1 // 32$) и $3^3 \times 4^1 \times 5^1 \times 8^1 \times 4^1 // 32$ (из плана $3^3 \times 4^1 \times 4^1 \times 8^1 \times 4^1 // 32$) для математического моделирования испытаний летательных аппаратов в естественных условиях. Для плана $3^3 \times 4^2 \times 9^1 \times 4^1 // 32$ среднее абсолютных величин парных коэффициентов корреляции коррелированных столбцов составляет $|\bar{r}_{ij}| = 0,066$, максимальная абсолютная величина $\max |r_{ij}| = 0,236$. Для плана $3^3 \times 4^1 \times 5^1 \times 8^1 \times 4^1 // 32$ $|\bar{r}_{ij}| = 0,107$, $\max |r_{ij}| = 0,392$. Все значения следует считать хорошими.

Для генерирования квазирегулярных квазиравномерных многофакторных планов экспериментов разработан алгоритм RASTA8 [4, с. 115–120]. В качестве исходного плана эксперимента используются ЛП $_{\tau}$ равномерно распределенные последовательности [5].

Эвристика генерирования плана эксперимента заключается в отображении определенного подмножества точек по каждой ЛП $_{\tau}$ равномерно распределенной последовательности ξ_i в определенный уровень F_i фактора. Генерирование плана эксперимента основа-

но на гипотезе о равномерном распределении ЛП_τ последовательности в многомерном пространстве и равномерном распределении точек различных ξ_i друг относительно друга. Для непрерывных факторов значения уровней определяются по формуле

$$X_{iu} = X_{i\min} + \xi_{iu}(X_{i\max} - X_{i\min}),$$

где ξ_{iu} – значение ЛП_τ равномерно распределенной последовательности для i -го фактора и u -го опыта; $1 \leq i \leq k$; $1 \leq u \leq N_{\text{ЛП}\tau}$; $0 < \xi_{iu} < 1$;

$X_{i\min}$, $X_{i\max}$ – минимальное и максимальное значения i -го фактора в эксперименте.

Для дискретных и качественных факторов интервал изменения ЛП_τ равномерно распределенных последовательностей $(0, 1)$ разобьем на s_i подынтервалов: $0, 1/s_i$; $1/s_i, 2/s_i$; ...; $(s_i - 1)/s_i, 1$. Каждому подынтервалу присвоим уровни $0, 1, \dots, s_i - 1$ фактора X_i . В матрице плана ЛП_τ равномерно распределенных последовательностей каждое значение ξ_{iu} заменим уровнем 0 или $1, \dots$, или $s_i - 1$ в зависимости от того, в какой подынтервал попало значение ξ_{iu} .

Проведенный вычислительный эксперимент по полученному плану $2^1 \times 3^2 \times 4^3 \times 5^1 \times 7^1 // 32$ дал следующие результаты: среднее значение абсолютных величин коэффициентов парной корреляции факторов $|\bar{r}_{ij}| = 0,06434$; максимальная абсолютная величина коэффициента парной корреляции факторов $\max|r_{ij}| = 0,2000$. Коэффициенты парной корреляции факторов F_i плана эксперимента приведены в табл. 1. Полученные результаты следует считать хорошими.

Таблица 1. Коэффициенты парной корреляции факторов F_i плана $2^1 \times 3^2 \times 4^3 \times 5^1 \times 7^1 // 32$

Факторы	F_1	F_2	F_3	F_4	F_5	F_6	F_7	F_8
F_1	1							
F_2	-0,03626	1						
F_3	-0,03626	0,0462	1					
F_4	-0,05929	-0,12293	-0,19382	1				
F_5	-0,08402	-0,01726	-0,01726	-0,03849	1			
F_6	-0,08402	0,051793	0,051793	-0,03849	0,2	1		
F_7	-0,02127	0,055062	0,027094	-0,09288	-0,03037	-0,03037	1	
F_8	-0,10636	-0,05955	0,132934	-0,06303	0,034841	0,034841	-0,03492	1

В реальных технических и технологических системах факторы могут быть коррелированы друг с другом. Одним из разработанных методов устойчивого оценивания статистических моделей является алгоритм RASTA13 [4, с. 187–189]. Он заключается в представлении уровней варьирования фактора, коррелированного с другими факторами, в виде двух факторов: физического и фиктивного. Оба фактора должны быть такими по числу уровней, чтобы с ними можно было оптимально планировать эксперимент.

Один из двух факторов может быть взят как реальный фактор, который имеет физический смысл и исследуется в эксперименте. Второй – как фиктивный (формальный) и не

имеющий физического смысла. Сочетание двух факторов позволяет функционально представить уровни коррелированного фактора в соответствии с требованиями матрицы плана эксперимента. В рабочей матрице эксперимента коррелированный фактор представляется заданными условиями значений уровней, которые он имеет в исходной постановке задачи.

В матрице плана эксперимента факторы представляются в виде физического X_i и фиктивного $X_{i\phi}$. В рабочей матрице проведения эксперимента коррелированный фактор X_j представляется своими уровнями.

Алгоритм RASTA13 был использован в плане эксперимента $3^1 \times 4^7 \times 7^1 \times 8^1 // 64$ при математическом моделировании конструкторских и технологических факторов, влияющих на прочность болтовых соединений из композиционных материалов типа углепластиков [6]. С основным (главным) фактором – номинальное значение диаметра болта $X_2(d) = 6; 8; 10; 12$ (мм) – коррелированы два фактора: толщина пластин болтового соединения с усилением $X_1(\delta_c)$ (мм) и количество прослоек усиления, распределенных равномерно по всей толщине пластины, $X_7(n)$ (шт).

$$r_{ij}(X_2(d), X_1(\delta_c)) = 0,6141,$$

$$r_{ij}(X_2(d), X_7(n)) = 0,4804.$$

Уровни варьирования коррелированных факторов $X_1(\delta_c)$ и $X_7(n)$ с основным (главным) фактором представляются следующими зависимостями:

$$X_1(\delta_c) = X_2(d) + [2 + (1/6)X_2(d)]X_{1\phi},$$

$$X_7(n) = 2 + [(X_2(d))/2 - 2]X_{7\phi}.$$

Для факторов $X_2(d)$, $X_{1\phi}$ и $X_2(d)$, $X_{7\phi}$ области совместного существования – прямоугольники, и коэффициенты парной корреляции между этими факторами равны нулю, так как факторы в парах ортогональны друг к другу.

По полученной многофакторной модели были найдены оптимальные значения конструкторских и технологических факторов и результаты использованы при производстве самолетов Авиационного научно-технического комплекса им. О.К. Антонова [6].

3. Эвристические решения в регрессионном анализе

Одна из проблем регрессионного анализа – выбор структуры многофакторной модели. В большинстве случаев структура модели исследователю заранее не известна.

За множество структурных элементов – главных эффектов и взаимодействий – принята структура полного факторного эксперимента. В полном факторном эксперименте все эффекты ортогональны друг к другу и их число равно числу опытов полного факторного эксперимента [4, с. 88].

В дробном факторном эксперименте при использовании многофакторного регулярного плана все главные эффекты ортогональны друг к другу. Если выбран план не близкий к насыщенному, то некоторые взаимодействия будут ортогональны к эффектам, введенным в структуру модели, или слабо коррелированы. Предполагается, что все эффекты нормированы. Тогда множество структурных элементов по количеству будет достаточным для адекватной аппроксимации результатов эксперимента, так как план по статистическим свойствам будет близким к плану полного факторного эксперимента.

Эвристичность предложенной структуры модели заключается в том, что любой дробный факторный эксперимент является определенной частью полного факторного экс-

перимента и структура модели для дробного плана является определенной частью структуры модели для полного факторного эксперимента.

Эвристический подход использовался при получении структуры модели в алгоритме RASTA3 [4, с. 81–82]. Число структурных составляющих N_{Π} , из которых выбирают необходимые элементы, всегда больше числа независимых результатов дробного факторного эксперимента $N_{\text{Д}}$. Выбор подмножества структурных составляющих для получения модели в общем случае не может быть выполнен однозначно. Однозначному выбору должно способствовать использование соответствующего дробного плана эксперимента и числа его опытов $N_{\text{Д}}$. Если эффекты не коррелированы, то для выбора структуры модели необходимо использовать алгоритм RASTA3 и программное средство «Планирование, регрессия и анализ моделей» (ПС ПРИАМ).

Принципиальное отличие выбора и введения в структуру модели структурных элементов по алгоритму RASTA3 от опубликованных пошаговых методов заключается в том, что элементы ортогональны или близки к ортогональным и, будучи введенными в модель, не выводятся из нее на последующих шагах формирования модели.

Примеры успешного использования алгоритма RASTA3 при решении реальных прикладных задач по сложным системам приведены в [4, с. 211–290; 7].

4. Выводы

Анализ примеров использования эвристики в планировании эксперимента и регрессионном анализе подтвердил гипотезу о целесообразности подхода в решении неструктурированных (качественно сформулированных) задач.

1. При построении многофакторных статистических моделей реальных сложных систем используются как формализованные математические решения, когда исходная информация о системе известна, так и эвристические неформализованные решения, если необходимая информация отсутствует. Последние получают исходя из содержательного анализа данных с учетом опыта работы исследователя в статистическом моделировании.

2. Эвристические методы должны обеспечивать устойчивое решение множества задач определенного класса. Правильность получения моделей оценивается по их критериям качества.

3. С использованием эвристических методов успешно реализован системный подход в получении многофакторных статистических моделей, который позволяет создавать надежную и устойчивую систему постановки эксперимента и обработки его результатов и получать корректное решение задач.

СПИСОК ЛИТЕРАТУРЫ

1. Лаборатория экспериментально-статистических методов исследований (ЛЭСМИ) [Электронный ресурс]. – Режим доступа: <http://www.n-t.org/sp/lesmi>.
2. Сайт кафедры «Технология машиностроения» Механико-машиностроительного института Национального технического университета Украины «Киевский политехнический институт» [Электронный ресурс]. – Режим доступа: <http://tm-mm.kpi.ua/index.php/ru/1/publications>.
3. Пушкин В.Н. Эвристика – наука о творческом мышлении / Пушкин В.Н. – М.: Политиздат, 1967. – 272 с.
4. Радченко С.Г. Методология регрессионного анализа: монография / Радченко С.Г. – К.: «Корнійчук», 2011. – 376 с.
5. Соболев И.М. Выбор оптимальных параметров в задачах со многими критериями / И.М. Соболев, Р.Б. Статников. – М.: Наука, 1981. – 111 с.
6. Математическое моделирование прочности болтовых соединений композиционных материалов типа углепластиков / С.Г. Радченко, С.Н. Лапач, А.З. Двейрин [и др.] // Открытые информационные

и компьютерные интегрированные технологии: сб. научн. тр. – Харьков: «ХАИ», 2014. – Вып. 63. – С. 61 – 71.

7. Радченко С.Г. Формализованные и эвристические решения в регрессионном анализе / Радченко С.Г. – К.: «Корнійчук», 2015. – 236 с.

Стаття надійшла до редакції 16.02.2015