УДК 519.237.5: 621.95 DOI: 10.20535/2409-7160.2023.XXIII.278760

Вибір структури регресійної моделі

Д.М. Кетько, С.М. Лапач
КПІ ім. Ігоря Сікорського, Київ, Україна

Анотація. *При побудові регресійних моделей конкретна структура її a priori в більшості випадків невідома. При цьому кількість можливих теоретичних варіантів моделей дуже велика. Однозначних теоретичних обґрунтувань вибору моделі не існує. Наявні рекомендації дуже часто при спробі практичного застосування протирічать одне одному. Пропонується при вибору маргінальної (конкретної) структури моделі опиратись на вимоги прикладної задачі, що можливо і при відсутності інформації про структуру. Пояснення приводяться на прикладі задач з дослідження стійкості свердел в залежності від їх геометричних параметрів. Розглядається прийняття рішення як про загальну структуру (вид формули), так і про конкретну (перелік членів моделі). Прийняття рішення опирається на мету побудови моделі і вимоги прикладної галузі до неї.*

Ключові слова: *регресійний аналіз, ортогональні поліноми Чебишева, стійкість свердел, маргінальна структура регресійного рівняння.*

Загальна структура моделі

Залежність стійкості свердла від величини заднього кута

Експериментальні дані (табл. 1) взяті з Степнова М.М. Для побудови емпіричних математичних моделей використовується одновимірний лінійний регресійний аналіз [1, 2, 3].

Таблиця 1.

Залежність стійкості від величини заднього кута

Задній кут (град)	6	7	8	9	10	11
Стійкість (хв.)	116	273	304	319	271	255

Оскільки з рис. 1 зрозуміло, що залежність нелінійна, будемо використовувати поліном другого порядку. Для визначення кращого варіанту опису будуємо дві різновидності поліноміальних моделей: звичайну і в ортогональних поліномах Чебишева. Моделі приведені нижче: Чебишева (1) і звичайний поліном (2).

$$\begin{aligned}&\hat{y} = 256,3333 + 50,28571 f_1^1(X) - 70,2976 f_1^2(\mathrm{X}),\\ &де\ f_1^1(X) = 0,4(X_1 - 8,5);\ f_1^2(X) = 1,875((f_1^1)^2 - 0,46667)\end{aligned} \tag{1}$$

$$\hat{y} = -1376,83 + 378,6321 X_1 - 21,0893 X_1^2 \tag{2}$$

Обидві моделі адекватні, інформативні. Показники інформативності і точності апроксимації (8,14%) обох моделей ідентичні.

Вільний член в моделі в поліномах Чебишева має фізичну сутність і дорівнює загальному середньому відгуку. В звичайному поліному вільний член є математичною абстракцією, не прив’язаною до фізичного процесу.

Аналіз стійкості (табл. 2) свідчить про ідеальну структурну і майже ідеальну обчислювальну стійкість моделі в поліномах Чебишева і незадовільні структурну і обчислювальну стійкість моделі в звичайних поліномах. Регресори в поліноміальній моделі мають взаємну кореляцію близьку до 1 і число обумовленості в мільйони одиниць, що робить неможливим обґрунтовано використовувати її.

Таблиця 2.

Аналіз стійкості моделі

Характеристика		Значення	
		Поліном	Поліном Чебишева
Структурна стійкість	Max(r_{ij})	0,9963	0
Обчислювальна стійкість	cond	5302883	1,7816

Якщо подивитись на табл. 3, в якій представлені статистичні характеристики, то можна побачити наслідки низької стійкості поліноміальної моделі при аналізі структурних відношень між регресорами (елементами моделі). Як можна бачити з таблиці, за поліноміальною моделлю більш вагомим є лінійна складова: всі її три характеристики вищі, ніж складової другого порядку. За моделлю Чебишева все навпаки. Простий погляд на графік дає зрозуміти, що більш вагомою є складова другого порядку, як і підтверджує Чебишевська модель. Крім того, значення таких показників, як β-коефіцієнт (занадто великі) і частка впливу (в сумі дає 46,92%, а не 89,76%) в поліноміальній моделі взагалі не заслуговує довіри. При значенні критерію Стюдента 3,18 в поліноміальній моделі обидва коефіцієнти значимі, тоді як за Чебишевською – лінійний член статистично не значимий.

Таблиця 3.

Статистичні характеристики регресорів

Регресор	Характеристика	Модель	
		Поліном	Поліном Чебишева
Першого порядку	t-розрахункове	4,52	2,80
	β-коефіцієнт	9,7508	0,5180
	Частка впливу (%)	26,83	26,83
Другого порядку	t-розрахункове	4,29	4,29
	β-коефіцієнт	-9,2668	-0,7933
	Частка впливу	20,09	62,93

Графіки експериментальної і модельної залежностей представлені на рис. 1

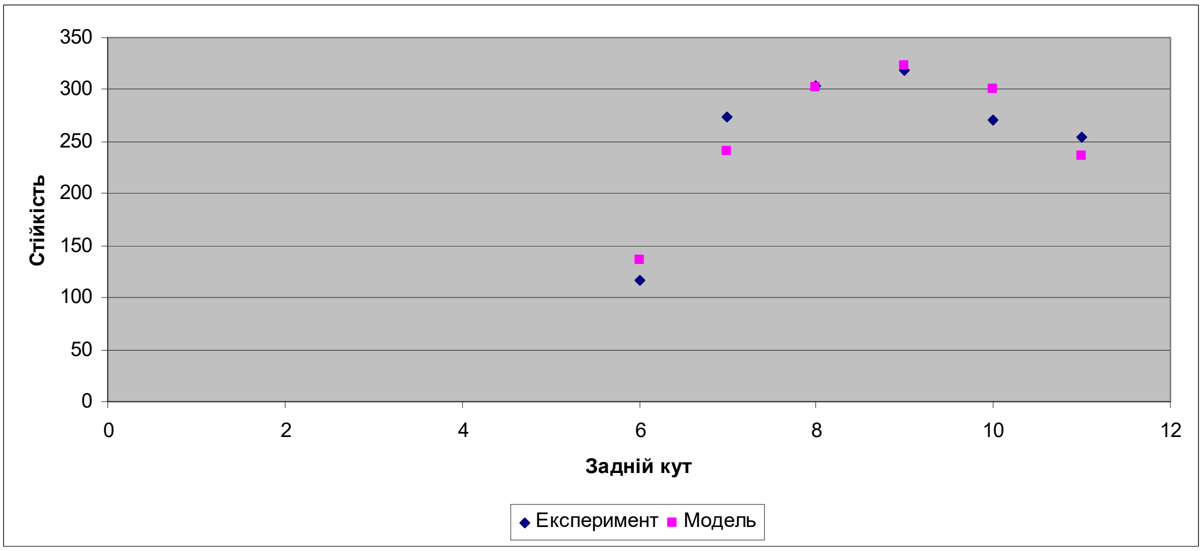

Рис. 1. Графік залежності стійкості від заднього кута

Висновок

Якщо потрібна тільки розрахункова формула, то в межах, яка завдається навчальною вибіркою, можна користуватись і звичайним поліномом. Якщо ж необхідний смисловий аналіз процесу, опираючись на модель, необхідно використовувати поліноми Чебишева.

Залежність стійкості свердла від зворотної конусності

Вихідні дані містяться в табл. 4. З рис. 2 видно, що ні лінійна, ні параболічна залежність не дає необхідного наближення даних. Необхідне перетворення вихідної змінної. Виходячи з форми залежності пропонується заміна $X' = {}^{1}\!/_{X}$ [4].

Таблиця 4.

Залежність стійкості від оберненої конусності

Зворотна конусність (мм)	0,01	0,02	0,03	0,04	0,05	0,06
Стійкість (хв.)	34,56	18,72	15,1	11,34	12,44	11,99

Від нової змінної отримані моделі першого і другого порядку, Модель другого порядку має наступний вигляд

$$\hat{y} = 17,35833 + 16,5661 f^1(X') + 1,417021 \mathrm{f}^2(\mathrm{X}'),$$

$$\text{де } f^1(X') = 0,016901(X' - 40,8333);$$

$$f^2(X') = 3,268697((f^1(X'))^2 - 0,62083 f^1(X') - 0,23375),$$

$$a\ X' = 1/X$$

Модель адекватна, інформативна, стійка і має добрі апроксимуючі характеристики.

Оскільки матриця поліномів Чебишева ортогональна, то для того, щоб з цієї моделі отримати модель першого порядку достатньо відкинути член другого порядку. Ніякі перерахунки моделі не потрібні (перерахунки характеристик моделі потрібні).

Характеристики стійкості моделей практично ідентичні. За характеристиками інформативності модель другого порядку незначно переважає модель першого.

З табл. 5 випливає, що член другого порядку статистично не значимий, але, його виключення різко знижує інформативність моделі, що свідчить на користь його присутності в моделі. Крім того, модель другого порядку забезпечує не тільки більш високу точність апроксимації, але й більшу її рівномірність на всьому інтервалі даних (табл. 6).

Таблиця 5.

Статистичні характеристики членів моделі

Коефіцієнти моделі		Характеристики			
Ім’я	Коефіцієнт	СКв помилка	t-розрахункове	β-коеф.	Частка впливу
$f^1(X)$	16,5661	0,886568	18,68565	0,990795	0,981674
$f^2(X)$	1,417021	0,755503	1,875601	0,099453	0,009891
Вільний член	17,35833		327,3651		
t критичне=	3,182446				

Таблиця 6.

Точність апроксимації

Середня абсолютна похибка апроксимації	0,98	0,55
Середня похибка апроксимації (%)	7,05	4,39
Максимальна абсолютна похибка апроксимації	1,40	1,60
Максимальна похибка апроксимації (%)	11,66	4,45

Таким чином, рекомендується вибрати модель другого порядку.

Список літератури

1. Norman R. Draper, Harry Smith Applied regression Analysis. Third edition / New York: Wiley-Interscience, 1998. 736 p.
2. С. Лапач, С. Радченко Математичне моделювання обробки високоміцних сталей // Mechanics and Advanced Technologies, 2019, т.85, №1, С.101–110.
3. Теорія планування експериментів: Виконання розрахунково-графічної роботи [Електронний ресурс] : навч. посіб. для студ. спеціальності 131 «Прикладна механіка», спеціалізації «Технологія машинобудування» / С.М. Лапач ; КПІ ім. Ігоря Сікорського. – Електронні текстові данні (1 файл: 3,31 Мбайт). – Київ : КПІ ім. Ігоря Сікорського, 2020. – 86 с. https://ela.kpi.ua/handle/123456789/38858
4. Лапач С.М. Лінійна регресія при прогнозування асимптотичних залежностей / Вестник Херсонского национального технического университета, №3(39), 2010г., –С.257–260.

Selection of the structure of the regression model

D.M. Ketko, S.M. Lapach

Abstract. *When constructing regression models, its specific structure is not known a priori in most cases. At the same time, the number of possible theoretical variants of the models is very large. There are no unequivocal theoretical justifications for choosing a model. The available recommendations very often contradict each other when trying to apply them practically. When choosing a marginal (specific) structure of the model, it is suggested to rely on the requirements of the applied problem, which is possible even in the absence of information about the structure.*
Explanations are given on the example of problems on the study of the stability of drills depending on their geometric parameters. Decision-making on both the general structure (the type of formula) and the specific one (the list of model members) is considered. Decision-making is based on the purpose of building the model and the requirements of the applied industry for it.
Keywords: *regression analysis, orthogonal Chebyshev polynomials, stability of drills, marginal structure of the regression equation*