УДК 519.254

МЕТОД ЗБІЛЬШЕННЯ ТОЧНОСТІ ПРОГНОЗНИХ РЕГРЕСІЙНИХ МОДЕЛЕЙ З МОЖЛИВІСТЮ ЗАСТОСУВАННЯ В СУЧАСНИХ КОМП'ЮТЕРНИХ ТЕХНОЛОГІЯХ

О.В. Ричка
Донецький національний технічний університет

У роботі запропоновано оригінальний метод підвищення точності регресійних прогнозних моделей, заснований на виключенні частини аномальних і малозначних вимірів поза "коридором". "Коридор" побудований на лінії рівняння $\hat{Y} = A \cdot x + B$ *на відстані 2k СКВ нев'язок* e_i*. Проведено порівняння з відомим методом Кука при використанні багатокритерійного підходу.*

В економічній і соціальній сферах, при здійсненні прогнозування, часто використовуються регресійні прогнозні моделі. Здатність регресійного рівняння відобразити взаємозв'язок між явищами знайшла собі практичне застосування в прогностичному аналізі. Однак існують дані, які становлять аномальні помилки. Це призводить до значного зниження якості прогнозування.

У реальних виробничих ситуаціях через вплив неврахованих у прогнозній моделі факторів помилок вимірів економічних характеристик, особливо при малих обсягах вибірки, точність прогнозу виходить досить низкою.

Метою досліджень є розробка нового методу підвищення точності прогнозних лінійних регресійних моделей, який буде досить простий у використанні й ефективний.

У даному дослідженні особливе місце займають порівняння запропонованого оригінального методу з методом-прототипом Кука за традиційними для регресійного аналізу критеріями та кількістю елементарних операцій ЕОМ для їх реалізації.

Сутність запропонованого методу підвищення якості лінійної регресійної прогнозної моделі полягає в наступному. На першому етапі дослідник, використовуючи всі вихідні статистичні дані, знаходить вид рівняння з використанням традиційного методу найменших квадратів $\hat{Y} = A \cdot x + B$. Далі визначаються відхили $e_i = Y_i - \hat{Y}$ та їх СКВ:

$$\sigma_e = \sqrt{\frac{\sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2}{n-2}}. \qquad (1)$$

Використовуючи (1) будуємо дві паралельні лінії, відстань між якими дорівнює $2k \cdot \sigma_e$ (2), де k – число (зазвичай $0,6 \leq k < 3$). Ці лінії відсікають із загального числа експериментальних даних як аномальні викиди, так і не досить вагомі для розглянутого регресійного рівняння виміри. Вага цих вимірів, що відкидаються, у величині коефіцієнта детермінації R^2 мізерно мала, але ці виміри істотно погіршують якість прогнозування.

При реалізації даного методу необхідно визначити межі, до яких варто зменшувати розміри "коридору", усередині якого вихідні статистичні дані вважаються надійними. Це завдання в розглянутому методі вирішуються шляхом визначення значень коефіцієнта детермінації R^2 при поетапному зменшенні величини k. При відкиданні частини статистики σ_e = varia.

Для порівняння методу- прототипу Кука й запропонованого методу використовувався відомий приклад [1] про взаємозв'язок віку дитини, у якому вона вимовив своє перше слово (X), і результати адаптивного тесту Геселля (Y).

На рис. 1 наведені дані цього прикладу. При використанні 100% вихідних даних лінійне регресійне рівняння має вид: $\hat{Y}=109,87 - 1,127 \cdot X$ ($R^2=0,41$). При $X_{прогн}=14,381$, $Y_{прогн}=93,67$. Довірчий прогнозний інтервал становить 4,8% від величини $Y_{прогн}=93,67$ при $P_{дов}=0,9$.

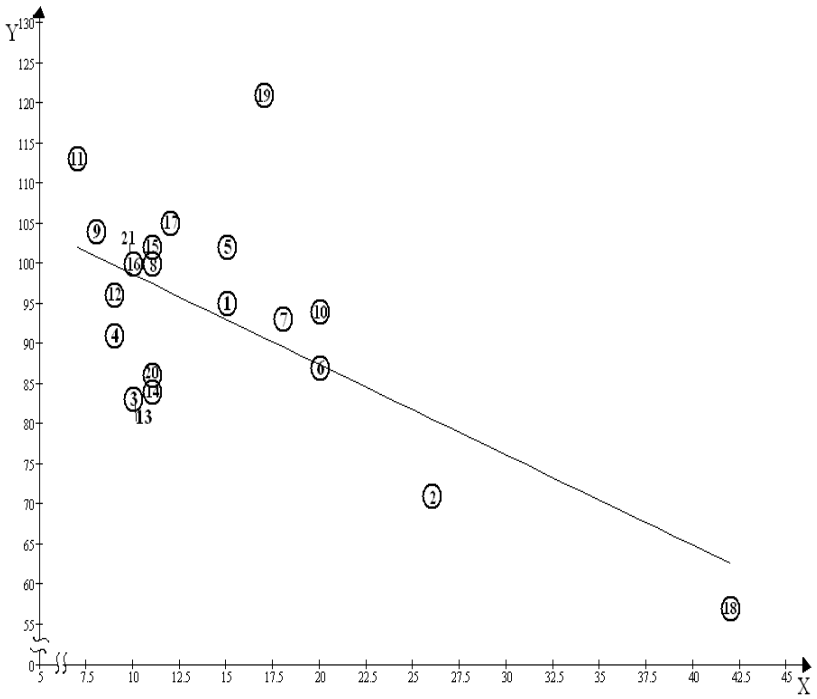

Рис. 1. Приклад регресійної моделі для тестування

Зсув $Y_{прогн}$ відсутній і Δ=0%. Для порівняння використовувалися наступні критерії ефективності:

− коефіцієнт детермінації R^2;

− модуль величини зсуву результату прогнозу (є наслідком зміни положення нового регресійного рівняння при відкиданні частини вихідних вимірів);

− довірчий інтервал прогнозних значень Yпрогн (являє собою геометричне місце розташування прогнозних значень Yпрогн при заданому значенні Хпрогн і заданої довірчої ймовірності Рдов) [2];

− кількість елементарних операцій ЕОМ, які необхідно для реалізації відкидання аномальних і ненадійних вимірів.

Результати порівняльного аналізу методу-прототипу Кука й нового оригінального методу наведені в табл. 1.

Таблиця 1

Результати порівняльного аналізу

Метод	Параметри методу	Номера виключених спостережень	Критерій якості методу			
			R^2	Δ^2, %	Величина довірчого інтервалу, %	Кількість елементарних операцій
Метод Кука	D_k=0,015	2, 3,19	0,58	0,32	~ 4	~ $0,4\cdot10^6$
	D_k=0,002	2, 3,19 11, 14,	0,64	0,18	3,8	~ $7\cdot10^6$
	D_k=0,097	2, 3, 14, 20, 19, 11, 5, 4	0,77	0,77	1,9	~ $102\cdot10^6$
Запропонований метод	k=1,27 (D_k=0,17)	19, 3, 13	0,7	0,12	2,4	~ $0,5\cdot10^3$
	k=1 (D_k=0,74)	3,13,14, 20, 19	0,83	1,96	2,2	~ $0,5\cdot10^3$
	k=0,8 (D_k=0,22)	2, 3, 13, 14, 20, 19, 11,5	0,86	1,35	1,8	~ $0,5\cdot10^3$

Висновки

1. Порівняння двох методів виявило, що метод- прототип Кука забезпечує виграш якості за рахунок більш ефективної стабілізації параметрів вихідного рівняння при відкиданні частини статистики, а запропонований метод – за рахунок підвищення значення R^2 і зменшення величини довірчого інтервалу прогнозу при заданій довірчій імовірності.

2. При досить малих і мало, відрізних між собою значеннях D_k можливі варіанти відкидання вихідної статистики, що істотно знижує величину R^2 регресійного рівняння після відкидання. Тому

рекомендується використовувати відомий метод Кука, тільки для ручного аналізу. У випадку регресійного аналізу на ЕОМ доцільно або повністю відмовитися від методу Кука або його використовувати разом із критерієм максимуму R^2 при відкиданні, без огляду на значний час аналізу.

3. Виграш у скороченні часу аналізу в порівнянні з методом-прототипом при $n \geq 20$ може скласти 10^5 та більше раз.

Бібліографічний список

1. Дрейпер Н.Р., Смит Г. Прикладной регрессионный анализ. 3-е изд.: Пер. с англ. − М.: Вильямс, 2007. − 912 с.
2. Справочник по специальным функциям с формулами, графиками и математическими таблицами./ Под. ред. М.Абрамовица и Н. Стигана: Пер. с англ. под ред. В.А. Диткина и Л.И. Кармазиной. − М.: Наука, 1979. − 830 с.
3. Кобзарь А.И. Прикладная математическая статистика. Для инженеров и научных работников. − М.: ФИЗМАТЛИТ, 2006. − 816 с.
4. Айвазян С.А. и др. Прикладная статистика: Основы моделирования и первичная обработка данных. Справочн. изд./ С.А. Айвазян, И.С. Енюков, Л.Д. Мещалкин. − М.: Финансы и статистика, 1983. − 471 с.
5. Справочник по прикладной статистике. В 2-х т. Т.1: Пер. с англ./ Под ред. Э. Ллойда, У. Ледермана, Ю.Н. Тюрина. − М.: Финансы и статистика, 1989. − 510 с.
6. Ханк Дж. Э, Райтс А. Дж., Уичерн Д.У. Бизнес-прогнозирование, 7-е изд.: Пер с англ. − М.: Вильямс, 2003. − 656 с.