



УДК 517.1:519.6

## ПРОГРАМНЕ ЗАБЕЗПЕЧЕННЯ ДЛЯ ПОБУДОВИ РЕГРЕСІЙНИХ ЗАЛЕЖНОСТЕЙ З ЗАСТОСУВАННЯМ «ФІКТИВНИХ» ЗМІННИХ

Студ. А. ООБурий., П. В0Влашук, гр. МгІТ-2-18,  
Козловський В. С., Сурнін А. А. МгЗІТ-18(л).  
Науковий керівник проф. С. М. Краснитський  
Київський національний університет технологій та дизайну

**Мета і завдання.** Мета полягає в розробці математичних та програмних компонентів побудови та дослідження регресійних моделей при застосуванні методу фіктивних змінних.

Завдання полягає в розробці програмного засобу для автоматизації процесу введення доцільної конструкції системи фіктивних змінних, котра знаходиться у відповідності із змістом задачі, що розв'язується.

**Об'єкт та предмет дослідження.** Об'єктом дослідження виступає процес побудови регресійних моделей з урахуванням якісного характеру частини параметрів моделі, а предметом дослідження виступає врахування зазначеного характеру шляхом введення так званих фіктивних змінних..

**Методи та засоби дослідження.** Теоретичною основою при вирішенні науково-прикладної проблеми є розробки закордонних та вітчизняних вчених в аспектах побудови математичних моделей природних, виробничих, соціальних та економічних процесів, математичного моделювання, математичного, програмного забезпечення САПР [2]. У теоретичних дослідженнях використано методи математичної статистики, теорії ймовірностей та прикладного регресійного аналізу. [1,2,3].

**Наукова новизна та практичне значення отриманих результатів.** На основі розробленого програмного забезпечення з'являється можливість ефективного комп'ютерного розв'язання задач побудови регресійних залежностей у випадку суттєвої ролі змінних якісного характеру і наглядного порівняння ефектів та наслідків від введення різних систем зазначених змінних.

**Результати дослідження.** Змінні, що розглядаються в регресійному аналізі, зазвичай можуть приймати значення з деякого неперервного діапазону. Проте інколи доводиться вводити фактори, що мають два або більше різних рівнів, причому такі рівні не мають ярко вираженого числового характеру. Наприклад, мова може йти про дані, що поступають з різних виробничих підприємств або від кількох операторів. У таких випадках не можна використовувати неперервну шкалу для змінних «підприємство», «станок», «оператор». Дослідник повинен надати таким змінним певні рівні, щоб можна було врахувати той факт, що різні підприємства, станки, оператори мають власні впливи на досліджуваний процес. Змінні такого типу мають назву *фіктивних*. Найпоширенішою фіктивною змінною є майже завжди застосована змінна  $x_0$ , що завжди дорівнює 1 і яка забезпечує наявність сталої величини  $\beta_0$  в залежності виду

$$y = \beta_0 x_0 + \beta_1 x_1 + \dots + \beta_{p-1} x_{p-1} + \varepsilon = \beta_0 + \beta_1 x_1 + \dots + \beta_{p-1} x_{p-1} + \varepsilon, \quad (1)$$

де  $x_1, \dots, x_{p-1}$  — регресори (незалежні змінні),  $\beta_1, \dots, \beta_{p-1}$  — коефіцієнти,  $y$  — відгук (залежна змінна),  $\varepsilon$  — випадкова складова.

Спосіб завдання фіктивних змінних не є унікальним. Зазвичай існує безліч способів завдання системи фіктивних змінних, що можуть відповідати будь-якій ситуації. Маючи в своєму розпорядженні певні «працюючі» вектори фіктивних змінних, можна одержати інші множини таких векторів за допомогою лінійних комбінацій векторів першої множини. Наступна множина обирається таким чином, щоб її вектори були лінійно незалежні від першої множини і т.д. В загальному випадку найбільш корисними вважаються прості множини фіктивних змінних, в яких використовуються рівні 0 та 1 або -1 та 1.

Наприклад, в моделі потрібно відобразити той факт, що певна продукція поставляється двома підприємствами А і В. Показники якості продукції є різними для згаданих підприємств. Для відображення зазначеного факту в модель може бути введена додаткова (фіктивна) змінна  $Z$  і відповідний коефіцієнт регресії  $\alpha$ , в результаті чого в моделі з'являється додатковий член  $\alpha Z$ , котрий повинен оцінюватися паралельно з іншими коефіцієнтами  $\beta_j$ . Змінна  $Z$  може приймати наступні значення:

$$Z = \begin{cases} 0, & \text{якщо продукція відноситься до підприємства А} \\ 1, & \text{якщо продукція відноситься до підприємства В} \end{cases}$$

Якщо  $a$  є оцінкою МНК (найменших квадратів) для  $\alpha$ , а  $f$  є іншою частиною моделі, то маємо залежність

$$\hat{y} = \hat{f} + \alpha Z. \quad (2)$$

Таким чином, дані для підприємства А оцінюються при  $Z = 0$ , і при цьому одержуємо

$$\hat{y} = \hat{f};$$

Оцінка для підприємства В відповідає значенню  $Z = 1$  і має вигляд

$$\hat{y} = \hat{f} + a.$$

Значення  $a$  оцінює різницю рівнів відгуків з групи даних підприємства В у порівнянні з відгуками з підприємства А з умови, що всі інші фактори представлені в  $\hat{f}$ .

Для визначення рівнів фіктивних змінних буває доцільним використовувати і інші значення  $Z$ . Припустимо, наприклад, що для підприємства А маємо  $n_1$  спостережень, а для підприємства В —  $n_2 = n - n_1$  спостережень. Наступні рівні змінної  $Z$ :

$$Z = \begin{cases} \frac{-n_2}{\sqrt{n_1 n_2 (n_1 + n_2)}} & \text{для підприємства А} \\ \frac{n_1}{\sqrt{n_1 n_2 (n_1 + n_2)}} & \text{для підприємства В} \end{cases}$$

забезпечують рівність 0 скалярного добутку стовпців, що відповідають змінним  $\beta_0$  і  $Z$  у регресійній матриці  $X$ . До того ж, стовець, що відповідає  $Z$ , буде нормалізованим, тобто сума квадратів його компонентів дорівнюватиме 1. Інколи такі властивості розширеної фіктивними змінними матриці  $X$ , відіграють значну роль в аналізі ситуації, а інколи можуть виявитися і несуттєвими. Вибір має зробити користувач-дослідник, і розроблене програмне забезпечення являє собою зручний засіб для обґрунтованого вибору.

За наявності кількох різних категорій даних, що використовуються, слід вводити додаткові фіктивні змінні. Найпростіший спосіб використовує наступну схему:

$$(Z_1, Z_2) = \begin{cases} (1; 0) & \text{для підприємства А} \\ (0; 1) & \text{для підприємства В} \\ (0; 0) & \text{для підприємства С} \end{cases}$$

У таких випадках також існує велика кількість можливих варіантів завдання належних рівнів, чому також сприяє розроблене програмне забезпечення.

**Висновки.** Розроблено програмне забезпечення для побудови і дослідження регресійних моделей при застосуванні методу фіктивних змінних [1].

Розроблений програмний засіб дає можливість ефективного введення різних систем фіктивних змінних і порівняння зазначених систем з точки зору більшої або меншої доцільності їх використання при розгляді конкретної задачі, що розв'язується.

**Ключові слова:** моделі регресії, фіктивні змінні, розділення блоків даних, ефекти взаємодії.

#### ЛІТЕРАТУРА

1. Прикладной регрессионный анализ: монография / Н.Р. Дрейпер, Г. Смит. — Москва – Санкт-Петербург – Киев, ДИАЛЕКТИКА.: 2017. – 911 с.
2. Щербань В.Ю. Математичні моделі в САПР. Обрані розділи та приклади застосування / В.Ю.Щербань, С.М.Краснитський, В.Г.Резанова – К.:КНУТД, 2010.-220 с.
3. Краснитський С.М. Векторні випадкові величини і випадкові процеси / С.М.Краснитський, В.Ю.Щербань та ін. – К.:Конус-Ю, 2008.- 191 с.