УДК 519.237.5

РОЗРОБКА ПРОГРАМНОГО ЗАБЕЗПЕЧЕННЯ ДЛЯ ДИСПЕРСІЙНОГО АНАЛІЗУ ТЕХНОЛОГІЧНИХ ПРОЦЕСІВ МЕТОДАМИ ЛІНІЙНОЇ РЕГРЕСІЇ

Студ. Д.В. Михайловський, гр. МгЗІТ-17(л)
Науковий керівник проф. С.М. Краснитський
Київський національний університет технологій та дизайну

Мета і завдання. Метою роботи є розробка програмного забезпечення для дисперсійного аналізу технологічних процесів методами лінійної регресії.

Об'єкт та предмет дослідження. Об'єктом дослідження є технологічні процеси різних видів. Предметом дослідження є дисперсійний аналіз технологічних процесів.

Методи та засоби дослідження. Дослідження ґрунтуються на основних положеннях математичної статистики і зокрема на методах регресійного та дисперсійного аналізу.

Для програмної реалізації розробленого алгоритму використовувалася середа C++ Visual Studio 2012.

Наукова новизна та практичне значення отриманих результатів. Так як технологічні процеси — це сукупність дій і операцій над початковими даними з метою отримання необхідного результату, то кількісний аналіз факторів, що впливають на течію конкретного процесу, дозволить налагодити більш продуктивний і оптимізований процес створення кінцевого продукту.

Результати дослідження. Насамперед, потрібно враховувати те, що технологічні процеси за своїм характером і призначенням є дуже різноманітними[1]. Наприклад, термічні процеси; барометричні процеси; каталізні процеси; електрохімічні процеси.

Дисперсійний аналіз (ДА) являє собою статистичний метод аналізу результатів будь-якого технологічного процесу на основі кількісного аналізу співвідношень, які залежать від якісних ознак. В будь-якому технологічному процесі середні значення досліджуваних величин змінюються у зв'язку зі зміною основних факторів (кількісних та якісних), що визначають умови процесу, а також і випадкових факторів. Дослідження впливу тих чи інших факторів на мінливість середніх є задачею дисперсійного аналізу.

Так, у моделі з фіксованим ефектами вважається, що виміри $y.$, котрі відповідають кожній комбінації рівнів $i_1 i_2 \dots i_m$ факторів $A_1, A_2, \dots, A_m$, можуть бути представлені сумою середнього, що відповідає комбінації цих рівнів, і похибки вимірювання. Тоді

$$y_{i_1 i_2 \dots i_m n} = \mu_{i_1 i_2 \dots i_m} + \varepsilon_{i_1 i_2 \dots i_m n}.$$

У випадку з двома факторами модель має вигляд

$$y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \varepsilon_{ijk}, \quad (1)$$
$$i = 1,\dots,I,\ j = 1,\dots,J,\ k = 1,\dots,K.$$

Параметри цього співвідношення мають такі назви: μ — загальне середнє, α_i — ефекти рядків, β_j — ефекти стовпців, $(\alpha\beta)_{ij}$ — ефект взаємодії, ε_{ijk} — випадкові складові.

Стандартні програми (ДА) обчислення оцінок генерального середнього і диференціальних ефектів, а також перевірки гіпотез щодо параметрів залежностей, мають певні обмеження в своїх застосуваннях, наприклад, передбачаючи обов'язково однакову кількість повторень експериментів при різних комбінаціях рівнів факторів або інші обмеження. Уникнути обмежень подібного роду дає можливість застосування методів лінійної регресії, що засновані на понятті загальної лінійної моделі і застосуванні методів перевірки загальної лінійної гіпотези. Про позитивні моменти вказаного застосування докладно сказано в спеціальній літературі [2-5].

У статистиці загальна лінійна модель регресії − це залежність між змінними y та X, що є лінійною за коефіцієнтами, тобто взаємозв'язок між даними моделюється за допомогою лінійних функцій, а невідомі параметри моделі оцінюються за вхідними даними. При використанні лінійної регресії взаємозв'язок між даними моделюється за допомогою лінійних функцій. Подібно до інших методів регресійного аналізу лінійна регресія повертає розподіл умовної імовірності y в залежності від X, а не розподіл спільної імовірності y та X, що стосується області мультиваріативного аналізу.

Загалом лінійна регресійна модель визначається у виді:

$$y = \beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k + u,$$

де y - залежна пояснювана змінна, $(x_1, x_2, \ldots, x_k)$ - незалежні пояснювані змінні, u - випадкова похибка, розподіл якої в загальному випадку залежить від незалежних змінних але математичне сподівання якої дорівнює нулеві.

Згідно з означенням моделі для кожного експериментального випадку залежність між змінними визначається формулами:

$$y_i = \beta_0 + \beta_1 x_{1,i} + \ldots + \beta_k x_{k,i} + u_i\,,\ i=1,\ldots,n.$$

або у матричних позначеннях $y = X\beta + u$;

Де:

$$y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}, \quad X = \begin{pmatrix} x'_1 \\ x'_2 \\ \vdots \\ x'_n \end{pmatrix} = \begin{pmatrix} 1 & x_{11} & \cdots & x_{1K} \\ 1 & x_{21} & \cdots & x_{2K} \\ \vdots & \ddots & \vdots & \\ 1 & x_{n1} & \cdots & x_{nK} \end{pmatrix}, \quad \beta = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_K \end{pmatrix}, \quad u = \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{pmatrix}$$

Зауважимо, що при застосуванні регресійного підходу до задач ДА матриця експерименту може опинитися погано зумовленою або навіть виродженою, що часто можна нейтралізувати належним введенням фіктивних змінних [4].

Висновки. Програмне забезпечення, що реалізує вищеописані кроки, дозволить покращити якість оцінювання результатів технологічних процесів та промислових експериментів за рахунок вищезазначеної модифікації відповідних процедур дисперсійного аналізу.

Ключові слова: *математична статистика, лінійна регресія, технологічні процеси, дисперсійний аналіз.*

ЛІТЕРАТУРА

1. Карташов М. В. Імовірність, процеси, статистика — Київ, ВПЦ Київський університет, 2007 — 75 с.

2. Афифи А., Эйзен С. Статистический анализ. Подход с использованием ЭВМ. — М:Мир, 1982. — 488 с.

3. Дрейпер Н., Смит Г. Прикладной регрессионный анализ. — М.: Диалектика, 2017.— 911 с.

4. George A.F. Seber and Alan J.Lee. Linear Regression Analiis. Second Edition. — Wiley-Interscience, 2013. — 856 p.