

УДК 519.21 + 681.3

РОЗРОБКА КОМП'ЮТЕРНОЇ ПРОГРАМИ ДЛЯ ДОСЛІДЖЕННЯ БАГАТОВИМІРНИХ СТАТИСТИЧНИХ ДАНИХ НА ОСНОВІ МОМЕНТНИХ ФУНКЦІЙ ПЕРШОГО ТА ДРУГОГО ПОРЯДКУ

Студ. О.Л. Мороз, гр. МгЗІТ-17(л)

Студ. Ніколаєв П.В., гр. МгІТ-3-17

Науковий керівник проф. С.М. Краснитський
Київський національний університет технологій та дизайну

Мета і завдання. Метою роботи є розробка комп'ютерної програми для дослідження багатовимірних статистичних даних на основі моментних функцій першого та другого порядку.

Об'єкт та предмет дослідження. Об'єктом дослідження є багатовимірні статистичні дані. Предметом дослідження є методи порівняння коваріаційних матриць багатовимірних вибіркового даних.

Методи та засоби дослідження. Дослідження ґрунтуються на основі оцінок векторних середніх значень і коваріаційних матриць вибірок з багатовимірних нормальних сукупностей, положеннях математичної статистики і комп'ютерній реалізації критеріїв Бартлетта з застосуванням апроксимацій Бокса F-розподілами та χ^2 – розподілом.

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

Наукова новизна та практичне значення отриманих результатів. Багатовимірний статистичний аналіз знаходить широке застосування в проектуванні і аналізі якості функціонування технологічних процесів, економічних дослідженнях і соціологічних дослідженнях. Тим не менш, завдяки складності виконання необхідних дій дослідники часто задовольняються застосуванням одновимірних моделей, що не можуть адекватним чином відтворювати відповідні процеси або явища. Тому є на часі розробка належних комп'ютерних програмних засобів, що з одного боку, були б нескладні для користувача-непрофесіонала, а з другого — забезпечували можливість реального використання методів і алгоритмів багатовимірного статистичного аналізу.

Результати дослідження. Практична реалізація методів і алгоритмів багатовимірної статистики для розв'язання конкретної проблеми може включати декілька типів одновимірного і багатовимірного аналізів з метою зрозуміти відношення між змінними та їх доцільність для фактичної проблеми, яку досліджують.

Деякі типи задач, які пов'язані з багатовимірними даними, наприклад проста лінійна регресія та множинна регресія, зазвичай не розглядаються як окремий випадок багатовимірної статистики, оскільки розглядаються з урахуванням (одновимірного) умовного розподілу однієї змінної, яка визначена іншими змінними. Як і в одновимірному, так і в багатовимірному випадку техніка обробки результатів спостережень багато в чому залежить від того, чи виконуються припущення про рівність дисперсій або коваріаційних матриць. Очевидний приклад цього дає дисперсійний аналіз. Саме тому доцільно проводити перевірку гіпотези про рівність дисперсій або коваріаційних матриць, перш ніж проводити дисперсійний аналіз. Припустимо, що ми маємо k незалежних вибірок обсягів n_1, n_2, \dots, n_k з p -вимірних нормальних сукупностей з середніми $\mu_1, \mu_2, \dots, \mu_k$ і коваріаційними $\Sigma_1, \Sigma_2, \dots, \Sigma_k$ матрицями відповідно. Далі, нехай потрібно перевірити гіпотези $H_0: \Sigma_1 = \Sigma_2 = \dots =$

\sum_k . Для перевірки можна скористатися одним узагальненням критерію Бартлетта заснованого на статистиці

$$M = (n - k) \ln |S_U| - \sum_{i=1}^k (n_i - 1) \ln |S_{U_i}|, \quad \sum_{i=1}^k n_i = n$$

Гіпотеза відкидається при великих значеннях статистики М. В силу того, що розподіл статистики М при справедливості гіпотези H_0 є досить складним, його апроксимують відомими розподілами.

1. Одновимірний критерій ($p=1$). В цьому випадку при справедливості гіпотези H_0 випадкова величина $v_2 M / [v_1 (b - M)]$ наближено розподілено як F_{v_1, v_2}

$$\frac{v_2 M}{v_1 (b - M)} \approx F_{v_1, v_2},$$

де

$$b = \frac{v_2}{1 - A_1 + 2/v_2}, \quad v_1 = k - 1, \quad v_2 = \frac{k+1}{A_1^2}, \quad A_1 = \frac{1}{3(k-1)} \left(\sum_{i=1}^k \frac{1}{n_i - 1} - \frac{1}{n - k} \right).$$

Ця апроксимація виключно точна.

2. Багатовимірний критерій ($p > 1$). При $p > 1$ нульовий розподіл статистики М апроксимується F-розподілом, основаним на співвідношенні

$$\frac{M}{C} \cong F_{v_1, v_2},$$

де
$$v_1 = \frac{(k-1)p(p+1)}{2}, \quad v_2 = \frac{v_1 + 2}{A_2 - A_1^2}, \quad b = \frac{v_1}{1 - A_1 - v_1/v_2}.$$

Висновки. Розроблено програмне забезпечення, що реалізує методи перевірки однорідності вибірок багатовимірних статистичних даних. Дане програмне забезпечення може бути застосовано для розв'язання часто виникаючих на практиці задач порівняння результатів різних технологічних процесів, опитувань різних верств населення, соціальних стандартів.

Ключові слова: статистичний аналіз даних, багатовимірні вибірки, моментні функції першого та другого порядку.

ЛІТЕРАТУРА

1. Мардиа К., Земроч П. Таблицы F-распределения — М.: Наука, 1984— 254 с..
2. Щербань В.Ю., Краснитський С.М. Резанова В.Г. Статистичні моделі в САПР — К.: КНУТД, 2011. — 220 с.
3. George A.F. Seber and Alan J.Lee. Linear Regression Analysis. Second Edition. — Wiley-Interscience, 2013. — 856 p.