УДК 517.1:519.6

# РОЗРОБКА ПРОГРАМНОГО ЗАБЕЗПЕЧЕННЯ ДЛЯ ПЕРЕВІРКИ ОДНОРІДНОСТІ БАГАТОВИМІРНИХ СТАТИСТИЧНИХ ДАНИХ

В.М. Григорян, магістрант
*Київський національний університет технологій та дизайну*
Н.Г. Сивун, магістрант
*Київський національний університет технологій та дизайну*
С.М. Краснитський, д. ф.-м. н., професор
*Київський національний університет технологій та дизайну*

Ключові слова: аномальні вибіркові дані векторного типу, матриці коваріацій, перевірка гіпотез про відсутність аномальності.

Сучасне програмне забезпечення, зокрема, реалізоване у виді комплексів прикладних статистичних програм, значну увагу приділяє питанням оцінювання параметрів і перевірки статистичних гіпотез у випадку вибіркових даних числового характеру. При цьому алгоритми аналогічних дій для векторних (багатовимірних) даних представлені у названих джерелах значно менш повно, і досить часто без належних роз'яснень з приводу використання можливостей, що мають місце. Зважуючи на сказане, автори статті розробили комп'ютерну програму, що розв'язує наступні задачі.

1) **Перевірка наявності аномальних векторних даних у даній вибірці**

Зазначена процедура заснована на обчислені так званої *вибіркової відстані Махаланобіса* $D^2$, явний вираз якої дається формулою (див., напр., [1,2]): $D^2 = (\boldsymbol{x} - \overline{\boldsymbol{x}})'\boldsymbol{S}^{-1}(\boldsymbol{x} - \overline{\boldsymbol{x}})$, в якій

$$\overline{\boldsymbol{x}} = \frac{1}{k}\sum_{i=1}^{k} \boldsymbol{x}_i, \qquad \boldsymbol{S} = \frac{1}{k-1}\sum_{i=1}^{k}(\boldsymbol{x}_i - \overline{\boldsymbol{x}})(\boldsymbol{x}_i - \overline{\boldsymbol{x}})'$$

$\boldsymbol{x}_i, i = 1, \dots, k$ — векторні вибіркові дані розмірності *p*, вже перевірені на наявність аномалій, $\boldsymbol{x}$ — вектор, що перевіряється, штрих — знак транспонування.. У випадку відсутності аномалій статистика $F = \frac{(k-p)k}{(k^2-1)p} D^2$ має розподіл Фішера $F_{p,k-p}$. При заданому α гіпотеза про аномальність $\boldsymbol{x}$ приймається при перевищенні статистикою $F$ квантиля рівня 1 – α зазначеного розподілу. Підкреслимо, що програма обчислює такі квантилі, і звертатися до додаткового забезпечення чи до статистичних таблиць немає потреби.

**2) Перевірка гіпотези про співпадіння коваріаційних матриць**

Нехай одержано $k$ незалежних вибірок об'ємів $n_{1,}$ $n_2, \dots, n_k$ з $p$-вимірних нормальних сукупностей з середніми $\mu_1, \mu_2, \dots, \mu_k$ і коваріаційними матрицями $\sum_1, \sum_2, \dots, \sum_k$ відповідно. Для багатьох статистичних процедур, зокрема для виконання дій дисперсійного аналізу, важливо знати, чи виконується гіпотеза $H_0: \sum_1 = \sum_2 = \dots = \boldsymbol{S}$. З цього приводу розроблена програма реалізує дії одного узагальненням критерію Бартлетта [2], заснованого на статистиці перевірки $M$, вираз якої є наступним: $M = (n-k)\ln|S_U| - \sum_{i=1}^{k}(n_i - 1)\ln|S_{U_i}|$, $\sum_{i=1}^{k} n_i = n$,

де $S_{U_i}$ — незміщена оцінка коваріаційної матриці $S_i$, що побудована за $i$-ю

$$S_U = \frac{1}{n-k} \cdot \sum_{i=1}^{k}(n_i - 1)S_{U_i}.$$

вибіркою,

Гіпотеза $H_0$ відкидається при великих значеннях статистики M. В силу того, що нульовий розподіл статистики M при справедливості гіпотези $H_0$ є досить складним, він апроксимується $F$-розподілом, заснованим на співвідношенні

$$Mb^{-1} \cong F_{v_1, v_2},$$

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

$$A_1 = \frac{2p^2 + 3p - 1}{6(k-1)(p+1)}\left(\sum_{i=1}^{k}\frac{1}{n_i - 1} - \frac{1}{n-k}\right), A_2 = \frac{(p-1)(p+2)}{6(k-1)}\left(\sum_{i=1}^{k}\frac{1}{(n_i-1)^2} - \frac{1}{(n-k)^2}\right)$$

Список використаних джерел

1. Афифи А., Эйзен С. Статистический анализ — М.: Мир, 1982 — 488 с.

2. Мардиа К. Таблицы $F$-распределения — / Мардиа К., Земроч З. М.: Наука, 1984 — 254 с.