2. http://www.ndu.edu.ua/liceum/web.pdf
3. https://webtune.com.ua/statti/web-rozrobka/struktura-sajtu/
4. https://atriples.com.ua/pravylna-struktura-saytu/

ХАЛАБУРСЬКИЙ В.В., КРАСНИТСЬКИЙ С.М.

КОМП'ЮТЕРНА ПРОГРАМА ДЛЯ ПЕРЕВІРКИ ГІПОТЕЗИ ПРО ОДНОРІДНІСТЬ БАГАТОВИМІРНИХ ПОКАЗНИКІВ ЕКОНОМІЧНИХ ТРЕНДІВ І ПРОЦЕСІВ

KHALABURSKY V.V., KRASNYTSKY S.M.

COMPUTER PROGRAM FOR VERIFYING THE HYPOTHESIS OF THE HOMOGENEITY OF MULTIDIMENSIONAL INDICATORS OF ECONOMIC TRENDS AND PROCESSES

Software has been developed that implements methods for checking the homogeneity of samples of multidimensional statistics. The calculation of the appropriate probabilistic characteristics of the distributions that act as test indicators is performed by the program itself. As a result, reference to special statistical programs or tables is not required. This software can be used to solve problems that often arise in practice, comparing the results of different technological processes, surveys of different segments of the population, social standards. Keywords: multidimensional data, dispersion matrices, Fisher distribution.

Вступ

Сучасне програмне забезпечення, зокрема, комплекси статистичних програм, багато уваги приділяє питанням оцінювання параметрів і перевірки статистичних гіпотез у випадку вибірок даних числового характеру. Питання виконання 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. Враховуючи сказане, автори статті розробили комп'ютерну програму, що розв'язує наступні задачі.

Основна частина

Комп'ютерна програма, про яку йде мова у вступі, може виконувати наступні дії.

1) **Перевірка присутності аномальних даних у векторній вибірці**

Зазначена процедура заснована на обчислені так званої *вибіркової відстані Махаланобіса* D^2, явний вираз якої дається формулою (див., напр., [1]):

$$D^2 = (\boldsymbol{x} - \overline{\boldsymbol{x}})' \boldsymbol{S}^{-1} (\boldsymbol{x} - \overline{\boldsymbol{x}}),$$

в якій

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

$\boldsymbol{x}_i, i = 1, \ldots, k$ — векторні вибіркові дані розмірності p, вже перевірені на наявність аномалій, $\boldsymbol{x}$ — вектор, що перевіряється. 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 рівність кількох коваріаційних матриць

Нехай одержано k незалежних вибірок об'ємів $n_1, n_2, \ldots, n_k$ з p-вимірних нормальних сукупностей з середніми $\mu_1, \mu_2, \ldots, \mu_k$ і коваріаційними матрицями $\sum_1, \sum_2, \ldots, \sum_k$ відповідно. Для багатьох статистичних процедур, зокрема для виконання дій дисперсійного аналізу, важливо знати, чи виконується гіпотеза $H_0: \sum_1 = \sum_2 = \ldots = \boldsymbol{S}$. З цього приводу розроблена програма реалізує дії одного узагальненням критерію Бартлетта [2], заснованого на статистиці перевірки M, вираз якої є наступним:

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

де S_{U_i} — незміщена оцінка коваріаційної матриці S_i, що побудована за i-ю вибіркою, $S_U = \frac{1}{n-k} \cdot \sum_{i=1}^{k}(n_i - 1)S_{U_i}$.

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, основаним на співвідношенні

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

де

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

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

Висновки

Розроблено програмне забезпечення, що реалізує методи перевірки однорідності вибірок багатовимірних статистичних даних. Обчислення належних імовірнісних характеристик розподілів, що виконують роль індикаторів перевірки, виконується самою програмою. Як наслідок, звертання до спеціальних статистичних програм або таблиць не потрібне. 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.

Література

1. "Statistical Tables for the Social, Biological, and Physical Sciences" by Albert A. Bennett and James O. Berger: Springer, 2021, 632 p.
2. "Tables of F-and Related Distributions with Algorithms" by Mardia K., Zemroch P.: Academic Press, 1978, 254 p.
3. Краснитський С.М. Щербань В.Ю. та ін. Векторні випадкові величини і випадкові процеси — К.: КНУТД, 2008. — 191 с.

ПУТІЄНКО В.Р., КРАСНИТСЬКИЙ С.М.

ЗАСТОСУВАННЯ ІМОВІРНІСНОГО КЛАСИФІКАТОРА БАЙЄСА ДЛЯ ВИЯВЛЕННЯ НЕСАНКЦІОНОВАНИХ ДОСТУПІВ ДО КОМП'ЮТЕРНИХ МЕРЕЖ

PUTIIENKO V.R., KRASNITSKIY S.M.

APPLICATION OF THE BAYESIAN PROBABILITY CLASSIFIER FOR DETECTING UNAUTHORIZED ACCESS TO COMPUTER NETWORKS

With the growing reliance on computer networks for critical information exchange and data storage, the need to ensure their security has become paramount. Unauthorized access to computer networks can lead to significant consequences, including data breaches, privacy violations, and financial losses. To address this issue, this paper explores the application of a Naive Bayesian classifier as a tool for detecting unauthorized access to computer networks. *Keywords: unauthorized access, computer networks, naive Bayesian classifier, intrusion detection*

Вступ

Виявлення несанкціонованого доступу до комп'ютерних мереж є складним завданням через постійну зміну природи кіберзагроз та досвідченості зловмисників. Для вирішення цього виклику звертають увагу на використання алгоритмів машинного навчання, оскільки вони ефективно аналізують великі обсяги даних мережевого трафіку і розпізнають шаблони, що вказують на спроби несанкціонованого доступу.