

## ІНТЕЛЕКТУАЛЬНИЙ АНАЛІЗ МЕДИЧНИХ ДАНИХ

Мунтян Ольга

Вінницький національний медичний університет імені М.І. Пирогова

### Анотація

*Дослідження присвячено проблемі розробки уніфікованої методики інтелектуального аналізу медичних даних в рамках нової методології, яка базується на системному аналізі та сучасній статистичній парадигмі. Розглядаються критерії узгодженості та їх реалізація в сучасних пакетах прикладних програм математико-статистичного аналізу.*

### Abstract

*The research is devoted to the problem of developing a unified methodology for the intellectual analysis of medical data in the framework of a new methodology based on system analysis and a modern statistical paradigm. Criteria of consistency and their realization in modern packages of applied programs of mathematical-statistical analysis are considered.*

Інтелектуальний аналіз медичних даних (ІАМД) відіграє надзвичайно важливу роль в доказовій медицині, а його методи є інструментом (засобом) як безпосереднього оброблення результатів клінічних досліджень, так і функціоналом інтелектуальних систем підтримки прийняття лікарських рішень. Одним із методів ІАМД є математико-статистичний аналіз (МСА) результатів клінічних досліджень, який дозволяє виявити закономірності в досліджуваних процесах і явищах, обчислити статистичні параметри вибірок та оцінити параметри генеральної сукупності, встановити і дослідити причинно-наслідкові взаємозв'язки між даними. Результати цього аналізу є підґрунтям для прийняття чи відхилення висунутих в дослідженні гіпотез, вироблення висновків й узагальнень, та за висловом відомого американського математика, одного з основоположників послідовного статистичного аналізу А. Вальда – «прийняття оптимальних рішень в умовах невизначеності».

Незважаючи на велику кількість наукових публікацій та навчально-методичної літератури, присвячених питанням методології МСА та методики застосування його процедур, проблема вибору методів МСА, адекватних задачам дослідження, та коректного їх застосування на сьогодні залишається актуальною.

В роботі [1] визначені завдання, які необхідно виконати для розробки уніфікованої методики МСА результатів наукових досліджень в рамках нової методології, яка базується на системному аналізі та сучасній статистичній парадигмі. Одним із завдань є визначення передумов застосування кожного методу МСА.

В цій роботі розглядаються методи аналізу виду розподілу значень ознак (даних експерименту) та їх реалізація в поширених пакетах прикладних програм (ППП) МСА даних. Такий аналіз має бути першим етапом МСА, оскільки від його результатів залежить вибір способу опису дескриптивних (описових) статистик, зокрема характеристик центральної тенденції і розсіяння, та вибір адекватних методів подальшого аналізу даних. Якщо спосіб (формат) опису дескриптивних статистик є принципово важливим для подання результатів в наукових публікаціях, то правильний вибір методів МСА, наприклад, для порівняння груп, має вирішальне значення.

Традиційно перевіряється відповідність розподілу значень кількісних ознак закону нормального розподілу в припущенні, що розподіл значень ознаки в генеральній сукупності визначається за вибіркою.

Таку перевірку можна виконати різними способами:

1-й спосіб (якісний): побудова гістограми розподілу значень ознаки і візуальна її оцінка – наскільки вона близька до дзвону нормального розподілу (зазвичай виділяється

на гістограмі червоною лінією). Гістограма – це стовпчаста діаграма, яка відображає частотні дані. Всі ППП МСА і табличні процесори офісних пакетів (MS Excel та LibreOffice Calc) включають засоби побудови гістограм.

2-й спосіб (кількісний): Обчислення дескриптивних статистик та їх порівняння.

Вказані статистики обчислюються за допомогою відповідних інструментальних засобів ППП МСА або за допомогою бібліотечних статистичних функцій табличних процесорів офісних пакетів: AVERAGE (середнє арифметичне), MODE (мода – найчастіше повторюване значення), MEDIAN (медіана – значення, розташоване посередині набору), KURT (ексцес), SKEW (асиметрія) тощо.

Відзначимо, що для нормального розподілу: Mean=Mode=Median, Skewness=0, KURT=3.

3-й спосіб (кількісний): оцінка симетричності розподілу ознаки з тільки додатними значеннями. Якщо  $SD < Mean/2$ , то розподіл можна вважати симетричним. Але при цьому неможливо оцінити ексцес.

4-й спосіб (надійний): перевірка статистичних гіпотез про вид розподілу.

Формулюються дві гіпотези:

– нульова гіпотеза  $H_0$  – про те, що розподіл ознаки в генеральній сукупності відповідає закону нормального розподілу (або близький до нього);

– альтернативна гіпотеза  $H_1$  – про те, що розподіл ознаки в генеральній сукупності не відповідає закону нормального розподілу.

Перевірка статистичних гіпотез про вид розподілу значень ознаки виконується за допомогою статистичних критеріїв узгодженості. Найчастіше застосовуються наступні критерії узгодженості:

– Колмогорова-Смірнова (статистика D) та з поправкою Лілліфорса для випадків, коли середнє значення (Mean) та середнє квадратичне відхилення (СКВ, Standard Deviation – SD) ознаки відомі апріорі невідомі, а обчислюються за вибіркою;

– Шапіро-Вілка (статистика W) – найбільш потужний, строгий та універсальний – для випадків, коли Mean та SD ознаки апріорі невідомі;

– Д'Агостіно (асиметрія, ексцес);

– Андерсона-Дарлінга (статистика A-Sq);

– Крамера-Мізеса (статистика W-Sq);

Якщо розподіл кількісних значень ознаки нормальний або близький до нього, то МСА, зокрема, порівняння груп за цією ознакою, виконується параметричними методами. Якщо ж розподіл кількісних значень ознаки відмінний від нормального або значення ознаки є категоріальними (якісними), то для МСА, зокрема, для порівняння груп за цією ознакою, слід застосовувати непараметричні методи.

В дослідженні детально розглядаються інструментальні засоби перевірки відповідності розподілу значень кількісних ознак закону нормального розподілу найпоширеніших ППП МСА IBM SPSS (Statistical Package for Social Science), Statistics (IBM Corp., США), Statistica (StatSoft Inc., США), TIBCO Statistica (TIBCO Software Inc., США), StatPlus та BioStat (AnalystSoft Inc., США).

### Список використаних джерел:

1. Мунтян О.А., Мунтян М.Л., Яровенко А.Г. Пакети прикладних програм статистичного аналізу результатів наукових досліджень / Суч. інф. технології та інноваційні методики навч. у підготовці фахівців: методологія, теорія, досвід, проблеми// Зб. наук. пр. – Випуск 49 / редкол. – Київ-Вінниця: ТОВ фірма «Планер», 2017. – 186 с. – С.135-138.