

УДК 004.4

к.ф.-м.н. Соломін А.В., Лучицький Р.Ю.,  
Чорний К.В., Шкурат О.С.,  
Національний технічний університет України  
„Київський політехнічний інститут”

## АВТОМАТИЗОВАНА СИСТЕМА СТАТИСТИЧНОЇ ОБРОБКИ ІНФОРМАЦІЇ

*Приведено обґрунтування актуальності проблеми удосконалення програмних засобів для забезпечення прикладних статистичних методів досліджень.*

*Запропонована реалізація в сучасному програмному середовищі інструментарію статистичної обробки інформації, що виконується в напівавтоматичному режимі та крім прискорення і збільшення ефективності статистичних досліджень сприяє також зменшенню вірогідності помилок, пов'язаних з неправомірним застосуванням відповідних статистичних методик.*

*Ключові слова: автоматизована система, статистичні засоби, статистичні критерії.*

**Актуальність роботи.** Статистичні методи обробки інформації набувають все більшого розповсюдження у всіх сферах діяльності і галузях господарства. Якщо раніше вони виступали головним чином інструментарієм наукових досліджень, то зараз відіграють все більшу роль при вирішенні інженерних, економічних, медичних, соціально-політичних задач, забезпечуючи при цьому реальний економічний ефект. Спектр сучасних застосувань статистичних засобів надзвичайно широкий – це, наприклад, і планування розміщення зупинок міського транспорту, мережі аптек, лікарень, шкіл, дошкільних закладів в місті, і прогнозування соціально-економічних характеристик, і оптимізація інформаційних, транспортних та пасажиропотоків, і планування вакцинації населення і т.п. Цей список нескінченний, оскільки практично всі сфери діяльності в реальних умовах для підвищення ефективності потребують використання статистичного інструментарію.

Проблема полягає в тому, що насправді статистика в математичному плані досить складна і потребує глибокої теоретичної підготовки. Це пов'язано з великою кількістю існуючих статистичних методів, особливостями їх застосування, а головне, з припущеннями і умовами, за яких ці методи можна використовувати. На останній фактор часто звертають не достатньо уваги, що призводить до помилкових наслідків, причому, іноді це лише економічні

втрати, а іноді – реальна небезпека, наприклад, у випадку неправомірних висновків досліджень щодо дієвості медичних препаратів.

Для допомоги в здійсненні статистичних розрахунків створено досить багато програмних засобів. Але всі вони потребують таки глибоких математичних знань щодо умов застосування відповідних статистичних методів.

**Метою роботи** є розробка програмного інструментарію для статистичних досліджень, що не вимагає від користувача глибокого знання математичних основ і сприяє виробленню вірних статистичних висновків та тлумаченню їх результатів.

Зазначений інструментарій створено у вигляді системи взаємодіючих програмних модулів у середовищі NI LabVIEW [1], яке зараз набуває якостей неформального стандарту в галузі приладобудування і прикладних досліджень та легко вбудовується в більшість сучасних програмно-апаратних комплексів. Система розроблених програмних модулів налаштована таким чином, що виконанню певного статистичного алгоритму передують автоматична перевірка відповідних умов, і всі ці модулі утворюють розгалужене дерево, через яке при виконанні програми проводиться користувач до кінцевого результату – статистично достовірного висновку.

Середовище NI LabVIEW (National Instrument Laboratory Virtual Instrument Engineering Workbench – середовище розробки лабораторних віртуальних приладів) дуже зручне для застосувань, що створюються та використовуються не програмістами, а спеціалістами інженерних галузей. Замість текстових мов програмування зі складними синтаксичними правилами тут використовується графічна мова G, що має більш звичний для інженерів вигляд блок-діаграм. Це середовище графічного програмування, яке широко використовується в промисловості, освіті, науково-дослідних лабораторіях та ін., як стандартний інструмент для збору даних, управління та тестування приладів, обладнання, технологічних ліній, для автоматизації різних видів експериментів, а також розширює функції і можливості звичайного програмування.

Додатковими перевагами середовища є наявність великої кількості вбудованих функцій і підпрограм, багатство розвинутих, зручних та наочних елементів для побудови інтерфейсів, можливість взаємодії з іншими середовищами і навіть створення відповідних веб-застосувань з віддаленим доступом до сервісів.

Розроблений програмний інструментарій для статистичних досліджень призначений для перевірки статистичних гіпотез про рівність середніх значень та про однаковість дисперсій двох вибірок, причому це реалізується в автоматичному режимі, тобто спочатку перевіряються передумови для

застосування якогось статистичного критерію, після чого або цей критерій застосовується, або відбувається перехід до іншого критерію [2].

Інтерфейс системи наведений на рис. 1.

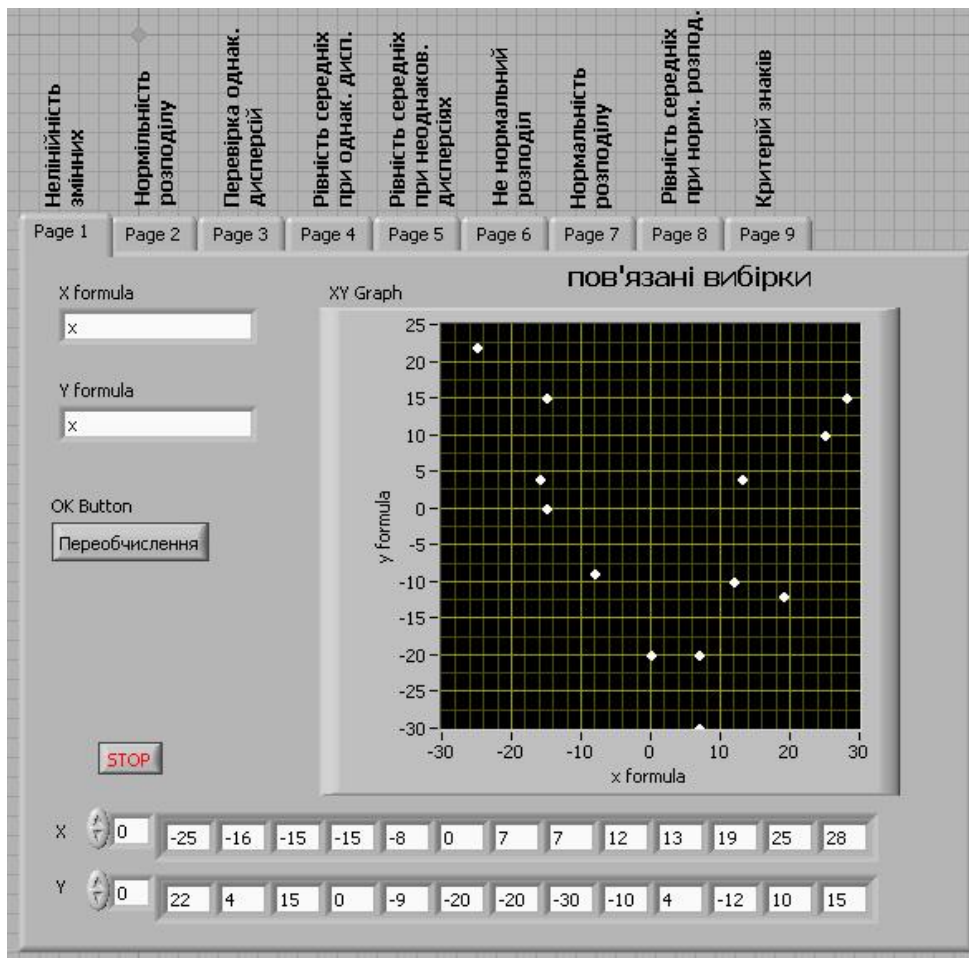


Рис. 1 – Інтерфейс програми

Алгоритм та програмний код можна умовно поділити на 9 основних модулів, що відповідають 9-м закладкам інтерфейсу, які перемикаються під час роботи програми синхронно з переходами між підпрограмами перевірки відповідних критеріїв та статистичних гіпотез.

Вхідна інформація вводиться в першому модулі у вигляді двох досліджуваних масивів в ручному режимі або у форматі файлів з таблицями, наприклад, Excel. Можливе пристосування до варіантів вводу з інших програмних засобів або вимірювальної апаратури. Далі пропонується обрати один із двох варіантів: вибірки містять попарно пов'язані дані або ні. У першому випадку відбувається перехід на сьомий модуль, інакше – на другий.

В другому модулі проводиться перевірка вигляду розподілу обох вхідних масивів (вбірок) на відповідність нормальному закону розподілу. Блок-діаграма підпрограми цього модулю наведена на рис. 2.

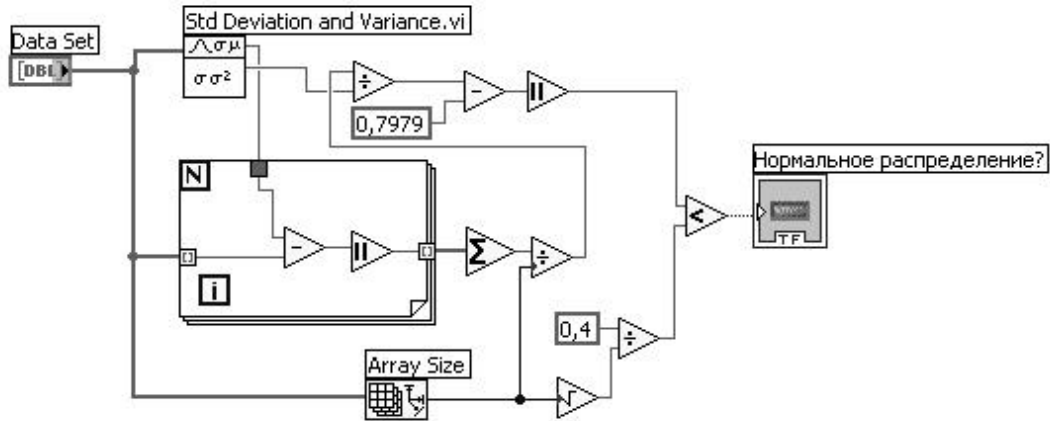


Рис. 2. Підпрограма перевірки вигляду розподілу вхідних масивів (вбірок)

В залежності від результату далі пропонується перехід на модуль 3 (перевірка однаковості дисперсій) чи на модуль 6 (застосування непараметричного критерію). Порівняння дисперсій в 3-му модулі виконується за допомогою критерію Фішера (реалізовано в підпрограмі з блок-діаграмою, зображеною на рис. 3).

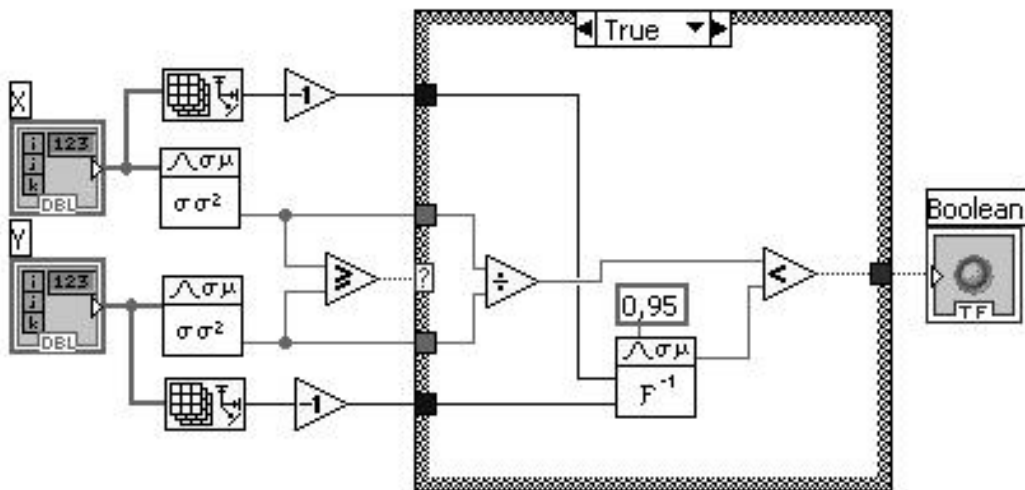


Рис. 3. Підпрограма порівняння дисперсій вибірок за критерієм Фішера

У випадку однаковості дисперсій пропонується перехід на модуль 4 (перевірка гіпотези про рівність середніх двох вибірок при однакових дисперсіях), інакше – на модуль 5 (перевірка гіпотези про рівність середніх двох вибірок при неоднакових дисперсіях).

Шостий модуль (рис. 4) проводить перевірку гіпотези про рівність середніх значень двох вибірок при ненормальному їх розподілі шляхом ранжування та застосування двовибіркового критерію Уїлкоксона. Використовуються розроблені підпрограми ранжування двох вибірок "Rank 2 arr" та перевірки критерію Уїлкоксона "W".

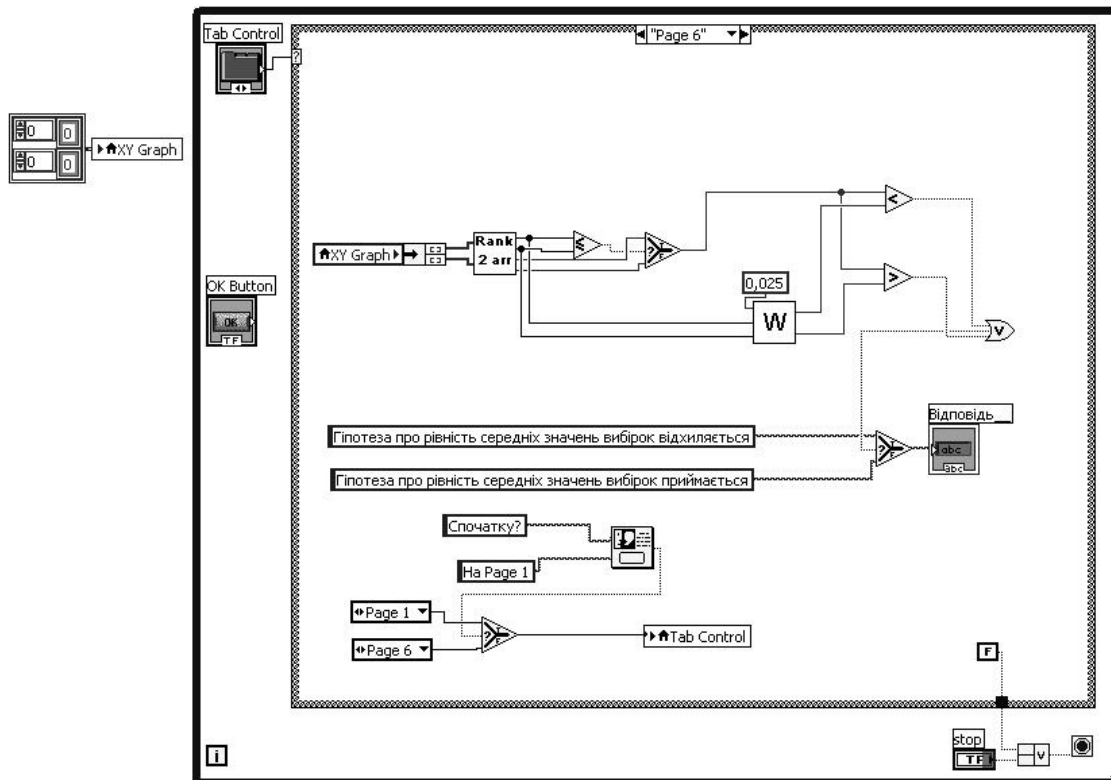


Рис. 4. Модуль перевірки гіпотези про рівність середніх за допомогою ранжування та критерію Уїлкоксона

Далі реалізуються статистичні алгоритми для пов'язаних вибірок, тобто коли досліджувані вибірки природним чином розбиваються на пари, наприклад, показники одних і тих же хворих до та після лікування (такі дослідження можна назвати перевіркою дієвості обробки).

У сьомому модулі знову проводиться тестування на відповідність нормальному закону розподілу, і у випадку позитивного результату відбувається перехід на восьмий модуль, інакше – на дев'ятий. У восьмому модулі застосовується спеціальний t-критерій Стьюдента для пов'язаних вибірок (рис. 5), а в дев'ятому – критерій знаків (рис. 6).

**Висновки.** Використання автоматизованої системи для перевірки статистичних гіпотез крім збільшення продуктивності та ефективності відповідних досліджень зменшує вірогідність помилок, пов'язаних з неправомірним застосуванням статистичних критеріїв.



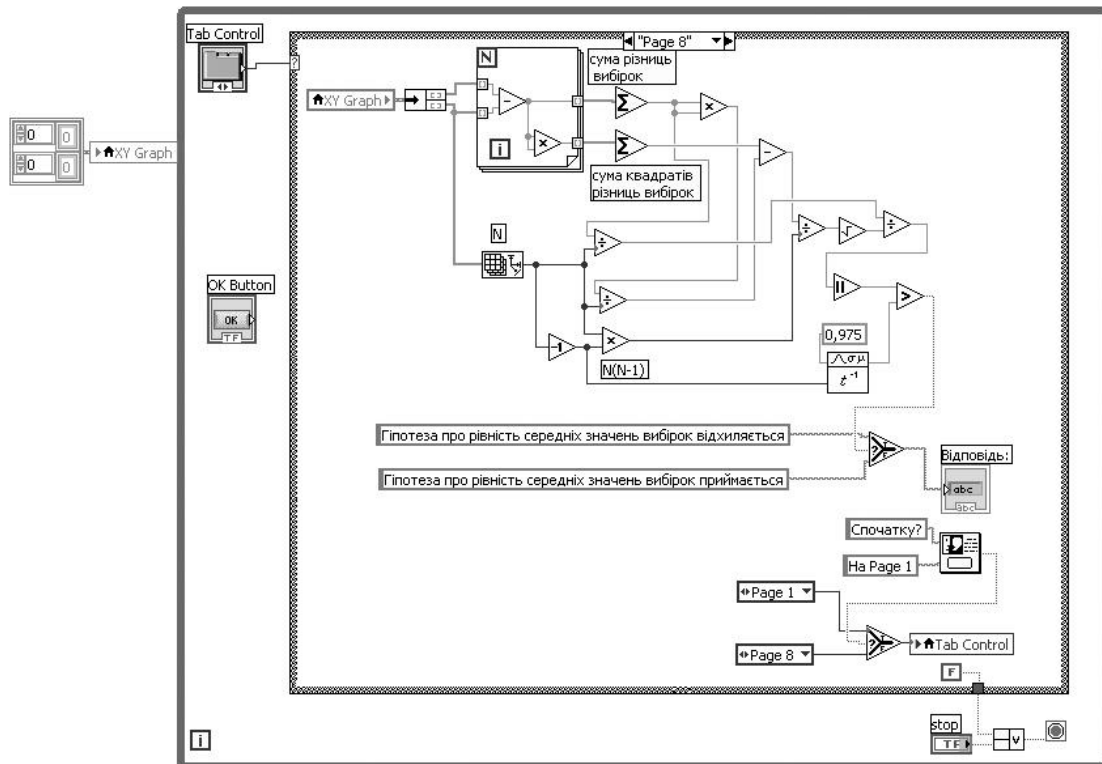


Рис. 5. Модуль перевірки гіпотези про рівність середніх значень для пов'язаних вибірок у випадку нормального закону розподілу

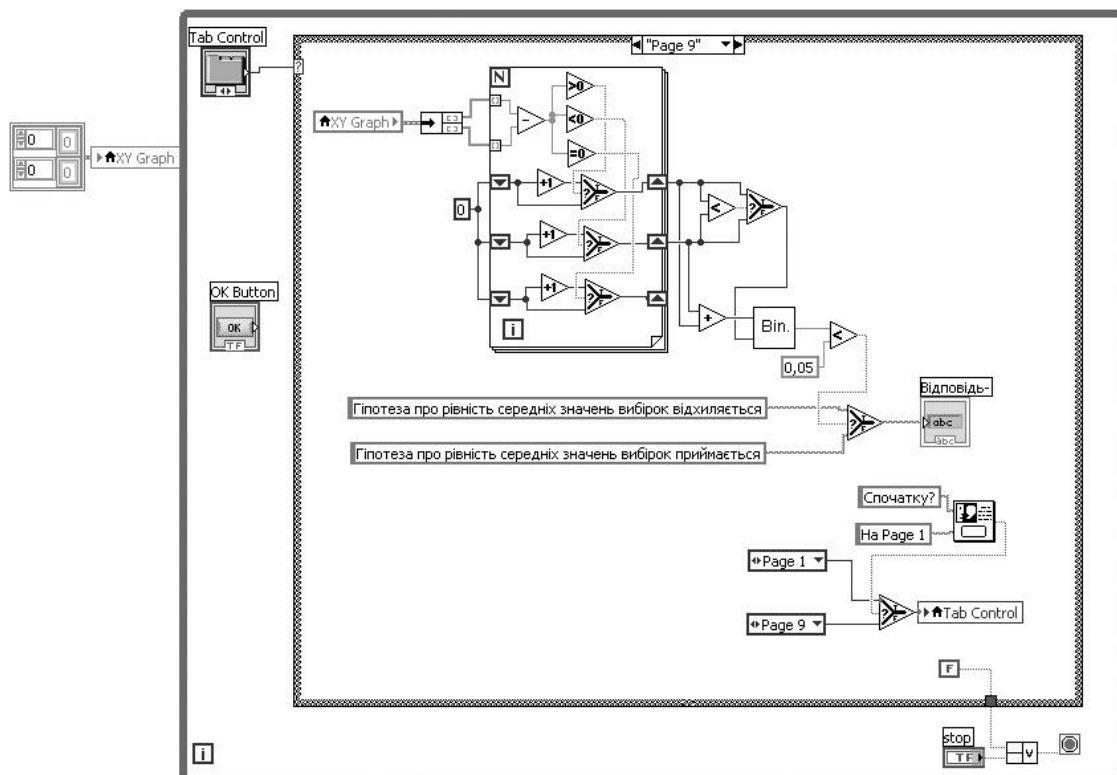


Рис. 6. Модуль перевірки гіпотези про рівність середніх значень для пов'язаних вибірок у випадку ненормального закону розподілу (за критерієм знаків)

Одним з варіантів використання системи є її розміщення на сервері у вигляді веб-застосування. У цьому випадку користувачам немає потреби встановлювати програму на власному комп'ютері, доступ до сервісу здійснюється через звичайний браузер.

Прикладів доцільності застосування системи у різних галузях можна навести безліч. Зокрема, це питання статистичного порівняння аналогічних регіонів, міст або районів міста для виявлення впливів окремих факторів; дослідження ступеню задоволення населення від втілення заходів удосконалення структури, наприклад, транспорту; аналіз результативності покращення екології, наприклад встановлення очисних споруд, на стан здоров'я місцевого населення.

### Література

1. Програмування в NI LabVIEW. Технологія розробки віртуальних приладів : навч. посіб. / О.Г. Кисельова, А.В. Соломін. – К. : НТУУ «КПІ», 2013. – 273 с.
2. Лапач С.Н. Статистические методы в медико-биологических исследованиях с использованием Excel / С.Н. Лапач, А.В. Чубенко, П.Н. Бабич. – К. : МОРИОН, 2001. – 408 с.

### Аннотация

Приведено обоснование актуальности проблемы усовершенствования программных средств для обеспечения прикладных статистических методов исследований.

Предложена реализация в современной программной среде инструментария статистической обработки информации, которая выполняется в полуавтоматическом режиме и кроме ускорения и увеличения эффективности статистических исследований способствует также уменьшению вероятности ошибок, связанных с неправомерным использованием соответствующих статистических методик.

Ключевые слова: автоматизированная система, статистические средства, статистические критерии.

### Annotation

The actuality of problem of improvement of programmatic tools for providing the applied statistical methods of researches is grounded.

Realization is offered in the modern software environment of tool for statistical treatment of information, which is executed in the semi-automatic mode and except for an acceleration and increase of efficiency of statistical researches promotes also to diminishing of probability of errors, related to illegal application of the proper statistical methods.

Keywords: CAS, statistical tools, statistical criteria.