

КОНЦЕПЦІЯ ІНФОРМАЦІЙНОЇ СИСТЕМИ ОБРОБКИ СТАТИСТИЧНИХ ДАНИХ НАУКОВИХ ДОСЛІДЖЕНЬ

Новожилова О. В., магістрант

Національний університет кораблебудування імені адмірала Макарова

Україна, Миколаїв

2151413@nuos.edu.ua

Анотація. Робота присвячена аналізу можливостей сучасних статистичних пакетів обробки даних та вирішенню науково-прикладного завдання створення інформаційного забезпечення обробки результатів статистичних даних, отриманих при проведенні наукових експериментів з подальшим використанням в науково-дослідницькій діяльності підприємств та учбових процесах.

Ключові слова: комп'ютерні статистичні пакети; статистичний аналіз даних; статистичні моделі; програмне забезпечення.

Вступна частина. Обробка статистичних даних полягає у визначенні значень числових характеристик, таких як математичне сподівання, дисперсія, асиметрія та ін. На основі отриманих значень характеристик формуються висновки щодо процесу дослідження. Такі розрахунки є досить трудомісткими і при виконанні вручну не гарантують необхідного рівня достовірності. Тому статистичну обробку дослідних даних доцільно виконувати за допомогою спеціалізованих прикладних програм для наукових розрахунків (MAXQDA, Mathcad 15.0, COSMOL Multiphysics, Wolfram Mathematica, Statgraphics, Statistica) [1,2].

Програмне забезпечення (ПЗ) для наукових розрахунків представлено багатьма організаціями: невеликими компаніями, корпораціями світового рівня. Спеціалізовані програми мають різне направлення - математика, астрономія, хімія, фізика, біологія, лінгвістика, інженерія, розробка штучного інтелекту тощо. ПЗ необхідне для відтворення графіків, обробки великих масивів інформації, зібраної науковими приладами. Згідно опитування SSI (Software Sustainability Institute), близько 92% вчених відповіли, що використовують у своїй практиці програми для наукових розрахунків. Більше половини з них розробляють власні програми, не маючи необхідних для цього навичок. А 69% опитаних висловили думку, що сучасні наукові дослідження просто неможливі без застосування комп'ютерної техніки і наукового софту [2].

У зв'язку з великим попитом статистичного аналізу даних практично в усіх сферах діяльності, особливо в науковій, ринок ПЗ для статистичної обробки даних швидко розвивається. Останнім часом отримали широке поширення програмні засоби або інформаційні системи, призначені для автоматизації робіт статистичної обробки даних, які дозволяють збирати, зберігати і обробляти різноманітні масиви даних з використанням єдиної інформаційної бази. Такі системи на підприємстві орієнтуються на потреби керівництва при виконанні функцій управління на основі внутрішніх і зовнішніх статистичних даних. Перевагами таких систем є адаптація інформаційної бази і функцій системи до умов функціонування підприємства. Проте, в силу складності реалізації математичних методів, такі системи, як правило, включають лише обмежений набір аналітичних методів.

Основна частина. При всьому багатстві програмного забезпечення статистичної обробки даних, етапи рішення статистичних завдань залишаються незмінними - побудова статистичної моделі, що відповідає завданню, вибір статистичного методу (разом з

програмою) і відповідного ПЗ, потім - реалізація цього методу в конкретному ПЗ для вирішення завдання. Отримання результату, що змістовно інтерпретується, не завжди означає закінчення дослідження. При використанні відібраних методів і засобів для аналізу інформації може бути потрібна безперервна робота впродовж декількох років. Використання ПЗ планується на усіх етапах дослідження, починаючи від організації зберігання даних і закінчуючи забезпеченням багатократного прогону завдань з різними параметрами.

Загальну логічну схему статистичного аналізу даних можна представити у вигляді семи етапів: початковий (попередній) аналіз досліджуваної системи; складання плану збору початкової інформації; збір початкових даних, їх підготовка і введення в ЕОМ; первинна статистична обробка даних; вибір основних методів і алгоритмів статистичної обробки даних, складання детального плану обчислювального аналізу матеріалу; реалізація плану обчислювального аналізу початкових даних (безпосередня експлуатація ЕОМ); підведення підсумків (будується формальний звіт про проведене дослідження, інтерпретуються результати застосування статистичних процедур) [3].

Проаналізувавши опубліковані результати досліджень, присвячених методам обробки статистичних даних, можна стверджувати, що більшість результатів вимірювання представляють собою вибірку з деякої генеральної сукупності. Особливо поширеними статистичними методами є: кореляційний аналіз, регресійний аналіз та факторний аналіз.[4]

Кореляційний аналіз. Між змінними (випадковими величинами) може існувати функціональний зв'язок, що проявляється в тому, що одна з них визначається як функція від іншої. Але між змінними може існувати і зв'язок іншого роду, що проявляється в тому, що одна з них реагує на зміну іншої зміною свого закону розподілу. Такий зв'язок називають стохастичним.

У регресійному аналізі моделюється взаємозв'язок однієї випадкової змінної від однієї або декількох інших випадкових змінних. При цьому перша змінна залежна, а інші - незалежні. Вибір або призначення залежної і незалежних змінних є довільним (умовним) і здійснюється дослідником залежно від вирішуваної їм задачі.

Факторний аналіз - один з найбільш популярних багатовимірних статистичних методів, він класифікує ознаки (змінні), що описують спостереження. Тому головна мета факторного аналізу - скорочення числа змінних на основі класифікації змінних і визначення структури взаємозв'язків між ними. Скорочення досягається шляхом виділення прихованих (латентних) загальних чинників, що пояснюють зв'язки між спостережуваними ознаками об'єкту, тобто замість початкового набору змінних з'явиться можливість аналізувати дані по виділених чинниках, число яких значно менше початкового числа взаємозв'язаних змінних.

Первинні дані, отримані при спостереженні, зазвичай важко інтерпретувати. Для того, щоб почати аналіз, в них потрібно внести деякий порядок і надати їм зручний для дослідника вигляд. Зокрема, спочатку бажано отримати уявлення про одновимірні розподіли випадкових величин, що входять в дані.

Візуалізація - це інструментарій, що дозволяє побачити кінцевий результат обчислень, організувати управління обчислювальним процесом і навіть повернутися назад до початкових даних, щоб визначити найбільш раціональний напрям подальшого руху.

Висновки. Проаналізувавши можливості та методи обробки статистичних даних із застосуванням спеціалізованих пакетів, можна зробити висновки, що вибір програмного забезпечення обробки статистичних даних - дуже складне питання. Для застосування ПЗ статистичних методів обробки даних користувачу не потрібно заглиблюватися в складність математичних процедур, а лише розуміти як ці методи використовуються. Для того, щоб

статистичний пакет програм був зручним та ефективним в роботі, він повинен задовольняти багатьом вимогам: містити значний спектр статистичного інструментарію (процедури, методи), відповідати вимогам введення, перетворення, збереження наборів даних, мати можливість експорту-імпорту даних, мати набір засобів візуалізації отриманих результатів, мати методичне забезпечення для самостійного освоєння користувачами, бути доступним за вартістю.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

- [1] Фетісов, В.С. (2018). *Пакет статистичного аналізу даних STATISTICA*. Ніжин: НДУ імені М. Гоголя.
- [2] *Програми для наукових розрахунків*. (н.д) <http://softlist.com.ua/catalog/programmy-dlya-nauchnykh-raschetov/> (18.10.2022).
- [3] Осипенко, Н.Б., & Осипенко, А.Н. (2015). *Программные средства первичной статистической обработки экспериментальных данных: практическое руководство*. Гомель: ГГУ им. Ф. Скорины.
- [4] Грицюк, П.М., & Остапчук, О.П. (2008). *Аналіз даних: Навчальний посібник*. Рівне: НУВГП.

Novozhilova O.

Concept of the information system for processing statistical data of scientific research

Abstract. *The work is devoted to the analysis of the capabilities of modern statistical data processing packages and the solution of the scientific and applied task of creating information support for processing the results of statistical data obtained during scientific experiments with further use in the scientific and research activities of enterprises and educational processes.*

Keywords: *computer statistical packages; statistical data analysis; statistical models; software.*