



УДК 519.21 + 681.3

РОЗРОБКА ПРОГРАМНОГО ТА МАТЕМАТИЧНОГО ЗАБЕЗПЕЧЕННЯ ДЛЯ ОБРОБКИ СТАТИСТИЧНИХ ДАНИХ ЗА ДОПОМОГОЮ ДІАГРАМ “СТЕБЛИНА З ЛИСТЯМ”

Студ. О.О. Іванов, гр. МгІТ-1-16
Науковий керівник проф. С.М. Краснитський
Київський національний університет технологій та дизайну

Мета і завдання. Створити програмне забезпечення для оцінки даних різного об'єму з можливістю отримання різних варіантів діаграми. Виходячи з того, що існуючі програмні пакети, такі як Statgraphics або функції Microsoft Excel, не дозволяють створити декілька варіантів діаграми з можливістю зміни даних, оптимальним рішенням є побудувати власну модель на мові Java задля вирішення таких задач швидко й з мінімальним процентом помилок у ході аналізу. В даному випадку мова програмування Java дає нам можливість реалізувати програмний код, який буде виконуватись на всіх пристроях, де є JVM(Java Virtual Machine) [1], з високою швидкістю.

Об'єкт та предмет дослідження. Об'єктом дослідження являються статистичні дані різного роду призначення. Предметом дослідження є аналіз статистичних даних за допомогою діаграм “Стеблина з листям”. Діаграма “стеблина з листям” являється представленням статистичних даних у вигляді описової статистики, яка дозволяє візуально оцінити форму, розмах і інші особливості розподілення досліджуваних даних. Загалом дані в діаграмі розташовані в відповідності з першими символами або стеблинами та замикаючими числами, або листям. Однією з переваг діаграми даного типу є те, що вона може візуалізувати великий масив інформації, суттєво збільшуючи можливості наглядного представлення характерних властивостей конкретного статистичного матеріалу [2].

Методи та засоби дослідження. Методом обробки статистичних даних є діаграми типу “Стеблина з листям”. Засобом дослідження є спроектоване та запрограмоване програмне забезпечення даного методу з великою кількістю налаштувань вибірки.

Наукова новизна та практичне значення отриманих результатів. Іноді недостатньо відділяти листя від стеблини по десятинним розрядам. При одному виборі може вийти так мало стеблин, що буде важко їх проаналізувати, через їх скупченість. Тому приходиться використовувати різні перетворення задля відділення листя від стеблини так як якість самої діаграми залежить від поданих статистичних даних. При сильній дисперсії даних, діаграма буде мати некоректний вигляд. Для покращення якості діаграми також можна переносити знаки, жертвуючи чіткістю даних. В такому випадку ми отримуємо якісну діаграму, яку вже можемо візуально оцінити. Також є можливість округлення чисел, зведення до другої, третьої степені, піднесення під корінь квадратний, та логарифмування вибірки. Наведемо приклад з книги [2] площини округів штату Міссісіпі.

Наприклад маємо вибірку: 448, 405, 729, 724, 412, 917, 592, ...

Округляємо її до десятків : 45, 40, 73, 72, 41, 92, 59, ... (#)

3* 8	(1)
4* 0121243121300214202	(19)
4* 597886556569	(12)
5* 142010	(6)
5* 977899958797	(12)
6* 412441	(6)
6* 898598	(6)
7* 320341203	(9)
7* 86657	(5)
8* 303	(3)
8* 8	(1)
9* 24	(2)

Оптимальний підбір доступних налаштувань вибірки може як покращити якість вибірки, так й понизити. Отже потрібна відповідальна оцінка даних.

Результати дослідження. Програмна модель створена на базі мови програмування Java, що дозволяє легко виконувати її на будь-якій платформі де є JVM. Обробка результатів експериментів виконувалася з застосуванням методів математичної статистики та теорії планування експерименту.

Висновки. На базі мови програмування Java розроблена програмна модель алгоритму, що реалізує процедуру представлення вибірових даних у формі «Стеблина з листям» при можливості створювати декілька різних варіантів зазначеної діаграми за допомогою маніпуляцій з даними вибірки, такими як та зміна системи утворення «стеблин» і «листя», а також застосування доцільних функціональних перетворень елементів вибірки. Стандартні програмні пакети не передбачають аналогічного варіювання відповідних представлень даних. В той же час реалізація вказаних можливостей суттєво розширює якість і швидкість розуміння характерних особливостей розподілу досліджуваних даних.

Ключові слова. Статистичний аналіз даних, діаграма «стеблина з листям», функціональні та просторові перетворення даних, бар'єрно-літерні представлення, програмне забезпечення, мова програмування Java.

ЛІТЕРАТУРА

1. Брюс Эккель. Философия Java 4-е издание. – СПб.: Питер, 2009. – 640 с.
2. Джон Тьюки. Анализ результатов наблюдений. Разведочный анализ. – М.: Мир, 1981. – 696 с.