Лабораторна робота

з дисципліни «Аналіз даних»

Тема: «Часткові та множинні коефіцієнти кореляції»

Виконав:

Студент групи ПМ-351

Подлєсний М.Д.

Перевірив:

завідувач кафедри

прикладної математики,

професор

Приставка П.О.

Зміст

[ПОСТАНОВКА ЗАДАЧІ 3](#__RefHeading__2814_198435239)

[ТЕОРЕТИЧНА ЧАСТИНА 4](#__RefHeading__2816_198435239)

[Часткові та множинні коефіцієнти кореляції. 4](#__RefHeading__2818_198435239)

[ПРАКТИЧНА ЧАСТИНА 7](#__RefHeading__2820_198435239)

[Записка користувача 7](#__RefHeading__2822_198435239)

[UML-діаграма 10](#__RefHeading__2824_198435239)

[ВИСНОВКИ 11](#__RefHeading__2826_198435239)

# ПОСТАНОВКА ЗАДАЧІ

На основі попередніх лабораторних робіт в рамках єдиної автоматизованої системи аналізу статистичних даних реалізувати такі обчислювальні процедури:

1. Знайти матрицю кореляції;
2. Матрицю частквої кореляції та оцінити її;
3. Підрахувати множинну кореляцію та переврити її на значущість.

# ТЕОРЕТИЧНА ЧАСТИНА

## Часткові та множинні коефіцієнти кореляції.

Оцінка часткового коефіцієнту кореляції визначає міру лінійної залежності поміж *і*-ої та *j*-ої ознаки, без урахування залежності цих двох змінних з елементами з набору *с.*

Обчислення оцінок часткових коефіцієнтів кореляціх виконують за рекурентними співвідношеннями. Якщо *і*,*j* і *d* – три різні змінні з множини *Х*, то усі оцінки часткових коефіцієнтів кореляції першого порядку визначають за виразом

,

де ,  та – значення оцінок парної кореляції для відповідних змінних.

Надалі послідовно застосовуючи рекурентну формулу

,

де *с* – будь-яка підмножина інших змінних.

Для перевірки на значущість часткового коефіцієнта кореляції



використовується статистика

,

де *w* – кількість змінних у наборі *с*. Статистика *t* має *t-*розподіл Стьюдента з *N-w-2*степенями вільності.

Оцінка множинного коефіцієнта кореляції rxk.x1,…xk-1,xk+1,…xn є мірою лінійної залежності поміж ознакою xk та набором x1,…xk-1,xk+1,…xn.

Оцінку коефіцієнта множинної кореляції зручно обчислювати за формулою

,

де





Для перевірки гіпотези



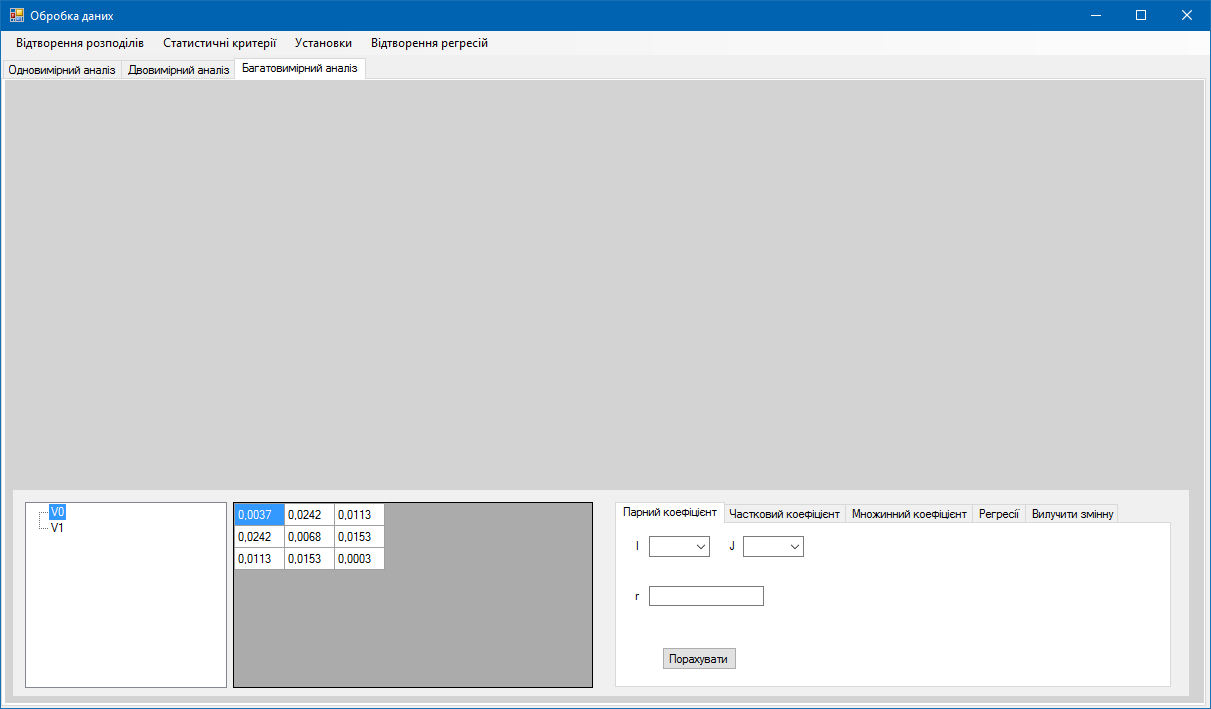
використовують статистику

,

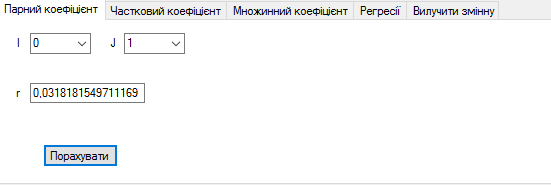
яка має *F*-розподіл з  і  степенями вільності.

# ПРАКТИЧНА ЧАСТИНА

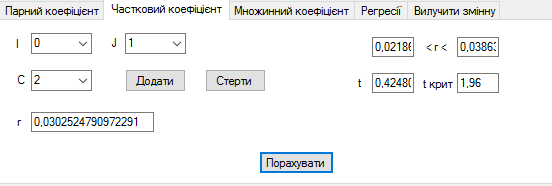
Обираємо вибірку з дерева



Обираємо ознаки для парного коефіцієнту кореляції

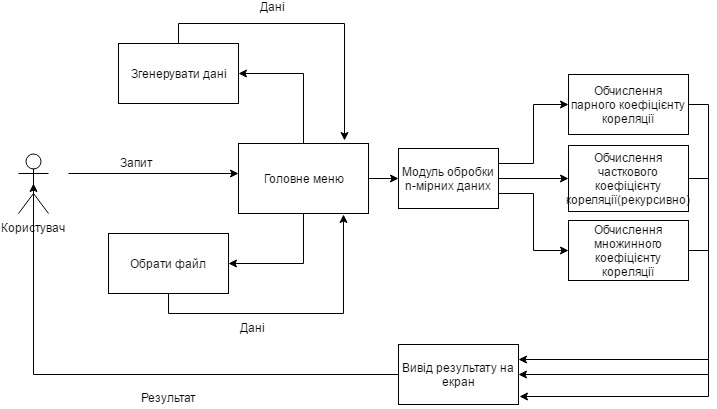


Те ж саме робимо для інших коефіцієнтів



Програма виводить оцінки, т-статистику та критичне значення

## UML-діаграма



# ВИСНОВКИ

На даній лабораторній роботі мною було реалізовано такі обчислювальна процедури:

1) перевірку наявності стохастичного зв’язку між ознаками об’єкту на основі часткових та множинних коефіцієнтів кореляції (за вимогою користувача):

* знаходження оцінок коефіцієнтів кореляції;
* перевірку їх значущості та для часткових коефіцієнтів кореляції призначення довірчих інтервалів (у випадку значущості);