**МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ**

**Національний авіаційний університет**

**Лабораторна робота № 2**

З дисципліни: «Аналіз даних»

Виконав:  
студент 351 групи ІІДС  
Подлєсний Максим

Перевірив:

завідувач кафедри  
Приставка П.О.

Київ-2016

**Зміст**

**Постановка задачі………………………………………………………..1**

**Теоретична частина……………………………………………………..2**

**Практична частина……………………………………………………..9**

**Висновки…………………………………………………………………12**

**Список літератури……………………………………………………..12**

**Постановка задачі**

На основі лабораторних робіт 1, 2 в рамках єдиної автоматизованої сис-

теми аналізу статистичних даних реалізувати такі обчислювальні процедури:

1) аналіз двовимірних об’єктів спостережень:

– проведення первинного статистичного аналізу двовимірних даних;

– відтворення двовимірного нормального розподілу;

– перевірку достовірності відтворення на основі критерію згоди χ2;

2) перевірку наявності стохастичного зв’язку між окремими ознаками

об’єкта:

– знаходження оцінки коефіцієнта кореляції, перевірку його значущос-

ті та призначення довірчого інтервалу (у випадку значущості);

– обчислення коефіцієнта кореляційного відношення та перевірку його

значущості;

**Теоретична частина**

**Обробка й аналіз двовимірних даних**

Розглянемо питання обробки та аналізу двовимірних масивів спостережень. Під час опрацювання таких масивів звичайно виникає три типи задач:

1) первинний аналіз, що включає побудову варіаційного ряду, перетворення даних, вилучення аномальних результатів спостережень, гістограмну

оцінку та перевірку нормальності розподілу двовимірної випадкової величини;

2) встановлення наявності стохастичного зв’язку між складовими двовимірного випадкового вектора;

3) за наявності стохастичного зв’язку між складовими випадкового вектора – задачі ідентифікації та відтворення регресії.

**Побудова варіаційного ряду**

Беручи за основу реалізацію ймовірнісної оцінки одновимірної випадкової величини, можна узагальнити подібну оцінку для випадку обробки масивів реалізацій двовимірних випадкових величин. Так, для реалізації

Ω2,N = {(xl, yl) l = 1…N}

двовимірного випадкового вектора

ζ = (ξ(ω),η(ω))

з функцією розподілу

F (x, y) = P{ω: −∞ < ξ(ω) < x,−∞ < η(ω) < y}

у припущенні незалежності складових ξ(ω) та η(ω)

F (x, y) = P{ω: −∞ < ξ(ω) < x}P{ω: −∞ < η(ω) < y}

можна розглядати одновимірні масиви

ξ(ω): xl; l = 1…N

η(ω): xl; l = 1…N

за кожним із яких можна провести побудову варіаційних рядів, розбитих на класи. Отже, визначаючи рівномірні розбиття Δhx, Δhy з кроками hx , hy відповідно за осями реалізацій величин ξ(ω) та η(ω) , автоматично задаємо рівномірне розбиття Δhx,hy площини реалізацій двовимірної випадкової величини площини реалізацій двовимірної випадкової величини ζ визначений за розбиттям Δhx,hy має такий алгоритм побудови.

1.За варіанту ряду {(xi, yj) i = 1…Mx, j = 1…My} де Mx , My – кількість елементів розбиття (класів) за відповідними осями, беруть центральну точку (i, j)-го елемента розбиття

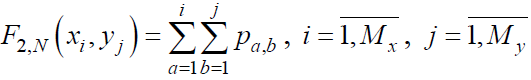
2. Із нижченаведених співвідношень визначають відносну частоту pij = nij/N

де ni, j – кількість точок вихідного масиву спостережень 2,N, що потрапили в

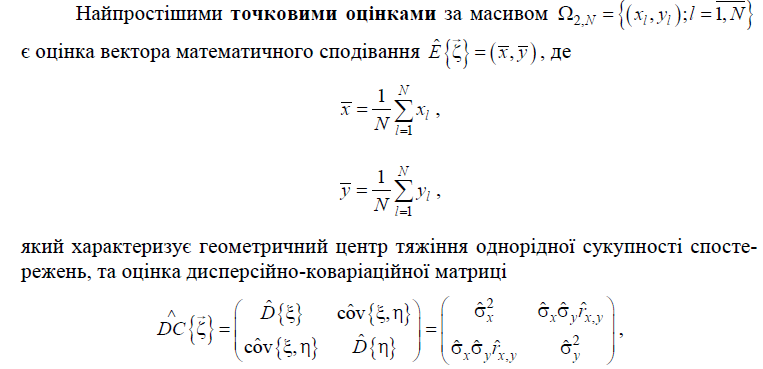
межі (i, j )-го елемента розбиття Δhx,hy. Як і у випадку одновимірних даних, відносні частоти з точністю доконстанти hx ⋅ hy є оцінкою усередненого значення функції щільності f (x, y)для неперервної випадкової величини ζ

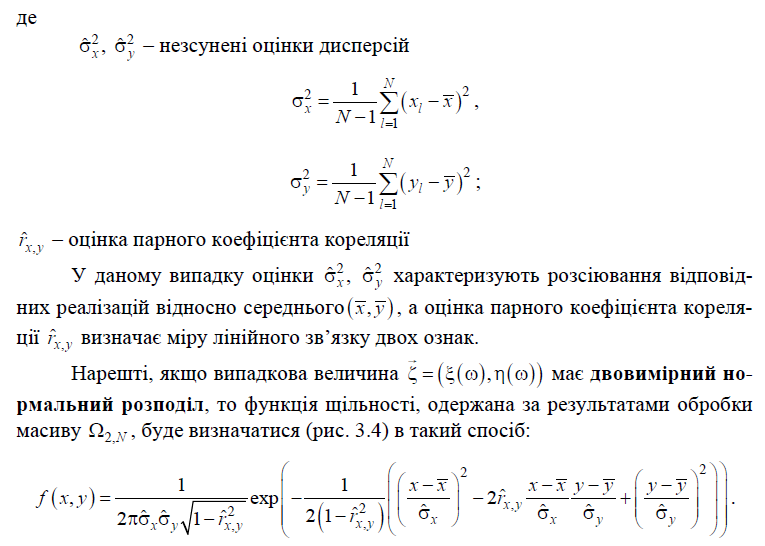
3. На основі відносних частот одержують емпіричну оцінку F2.N(x,y)

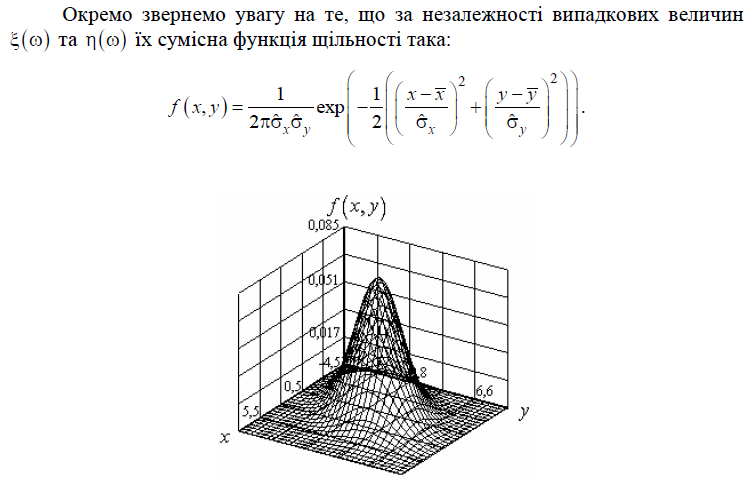
функції розподілу ζ

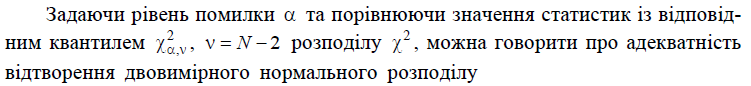
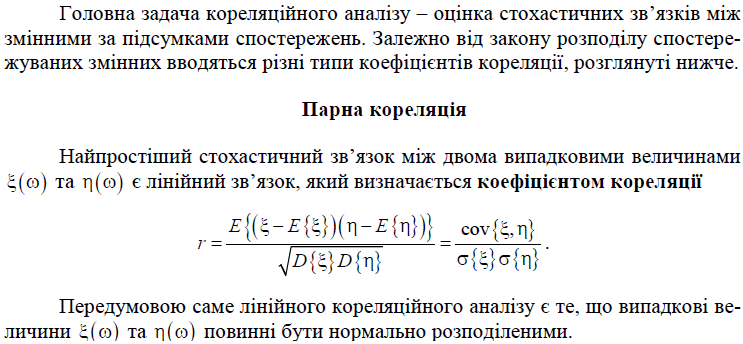
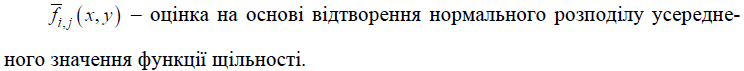
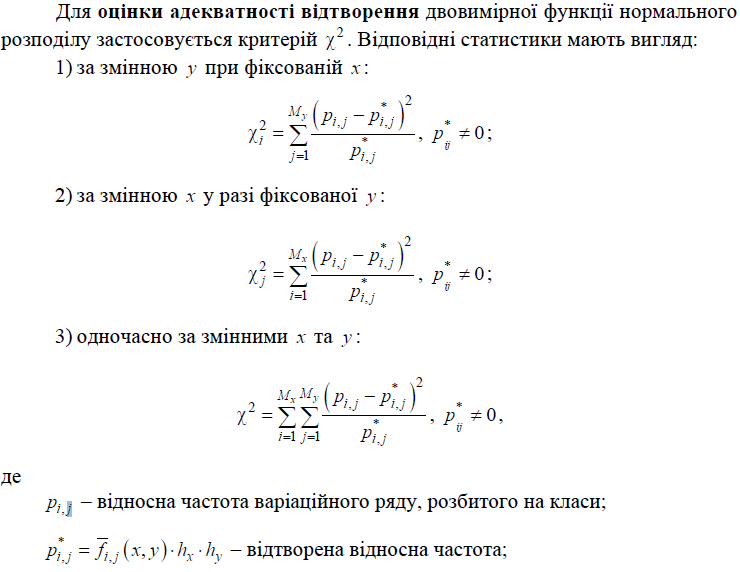


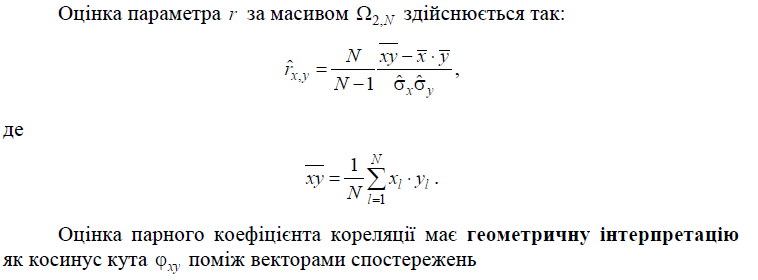
Побудований таким чином варіаційний ряд можна зобразити у вигляді двовимірної гістограми відносних частот

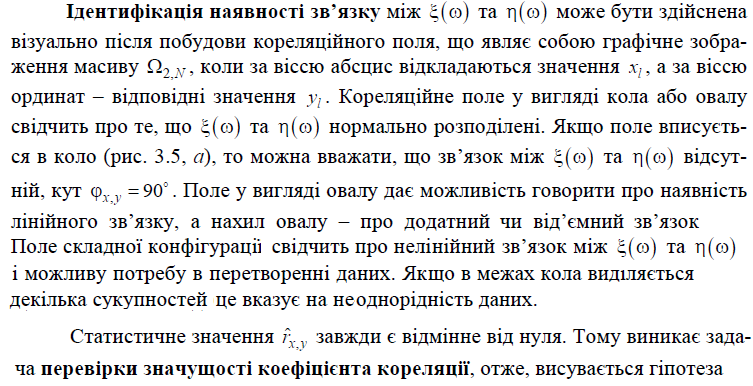
**Точкові оцінки**

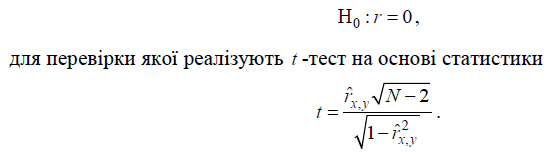


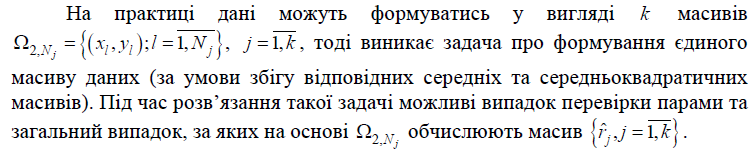


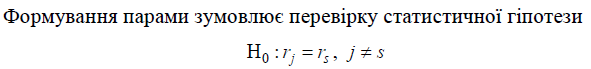
******Кореляційний аналіз**

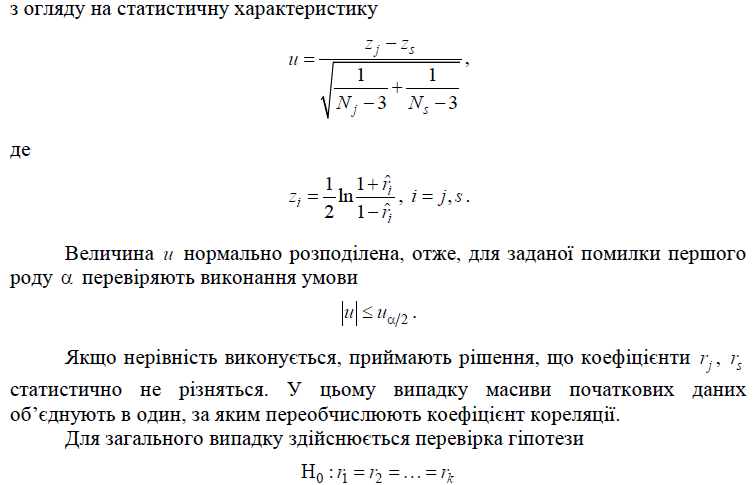
****

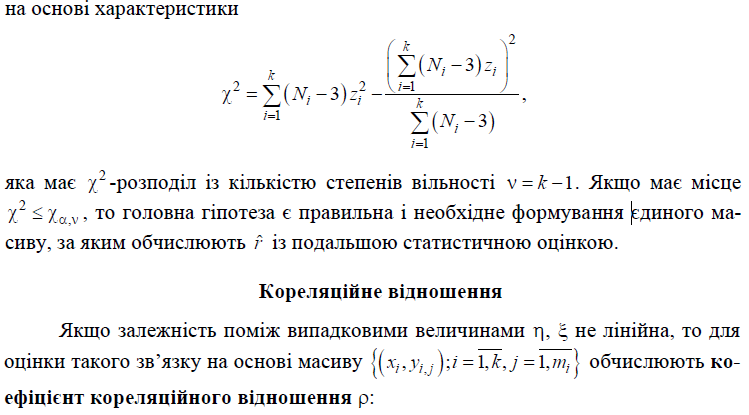


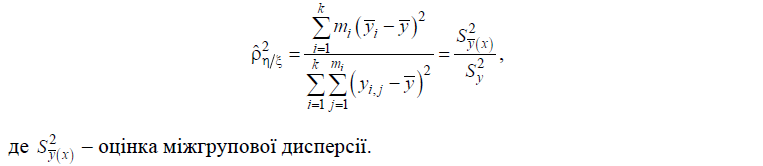


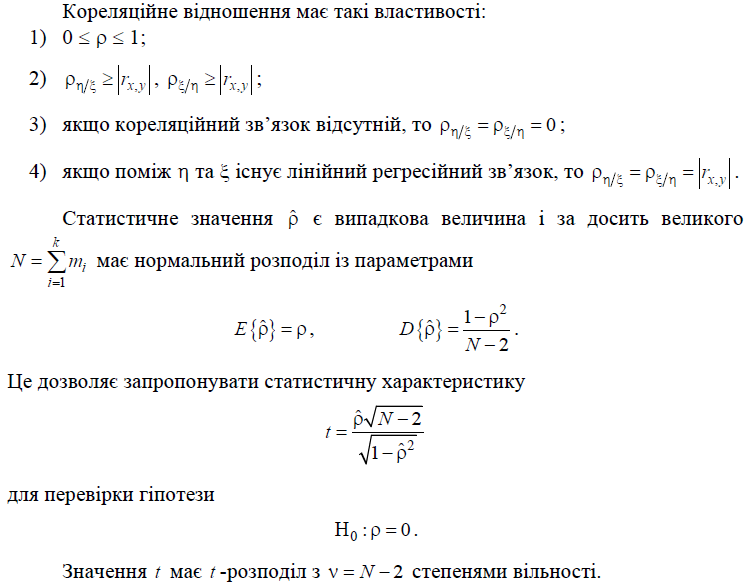


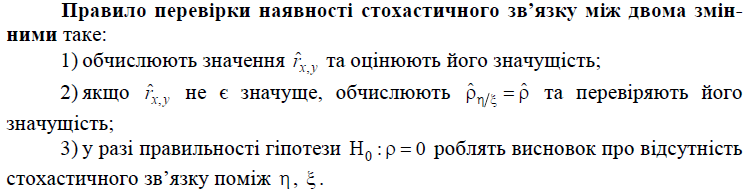




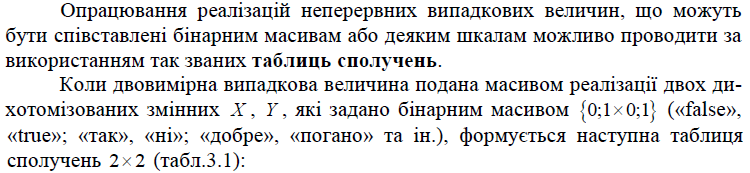




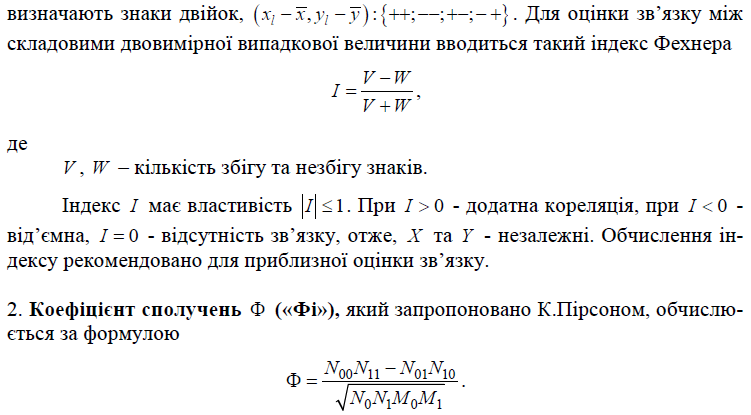


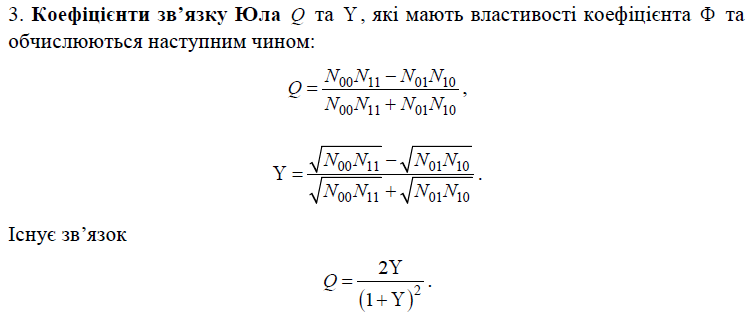


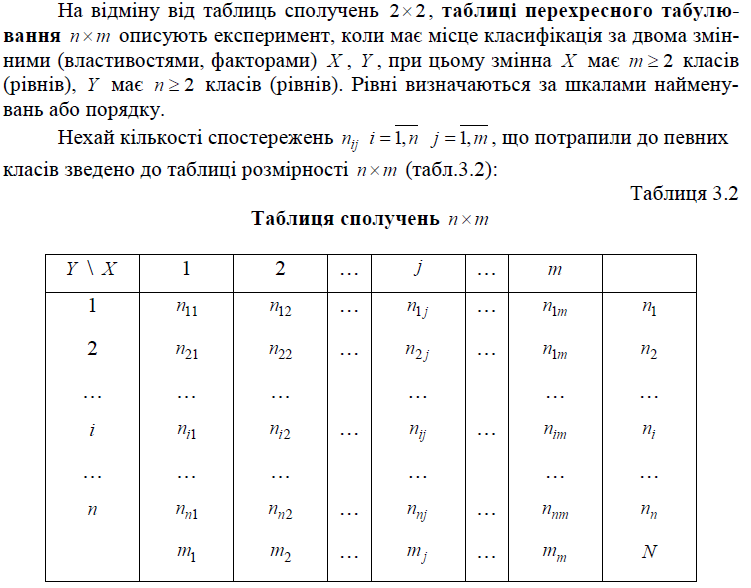
**Таблиці сполучень**

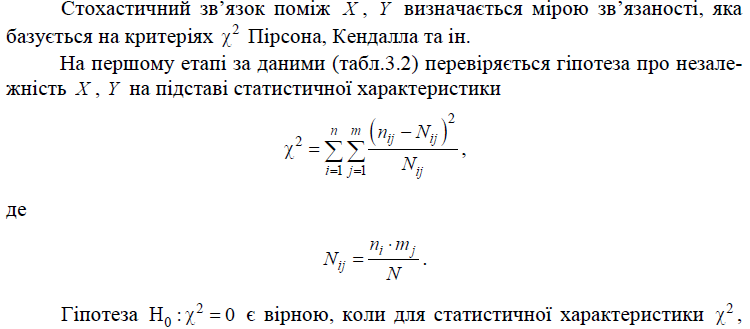
****

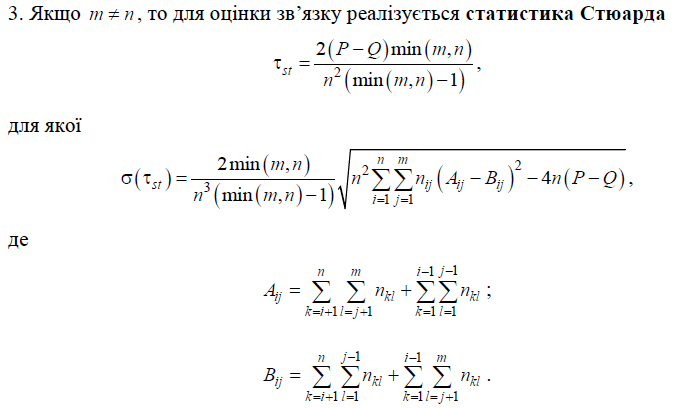
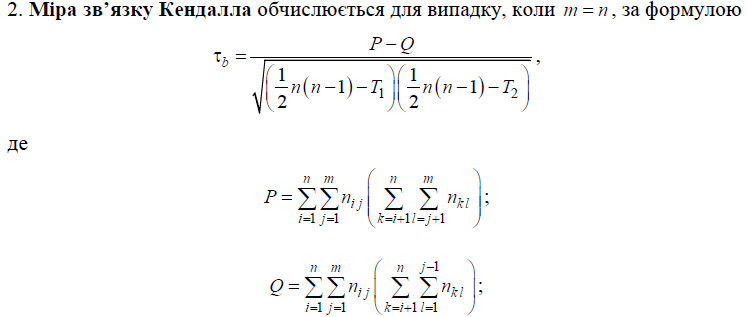
****



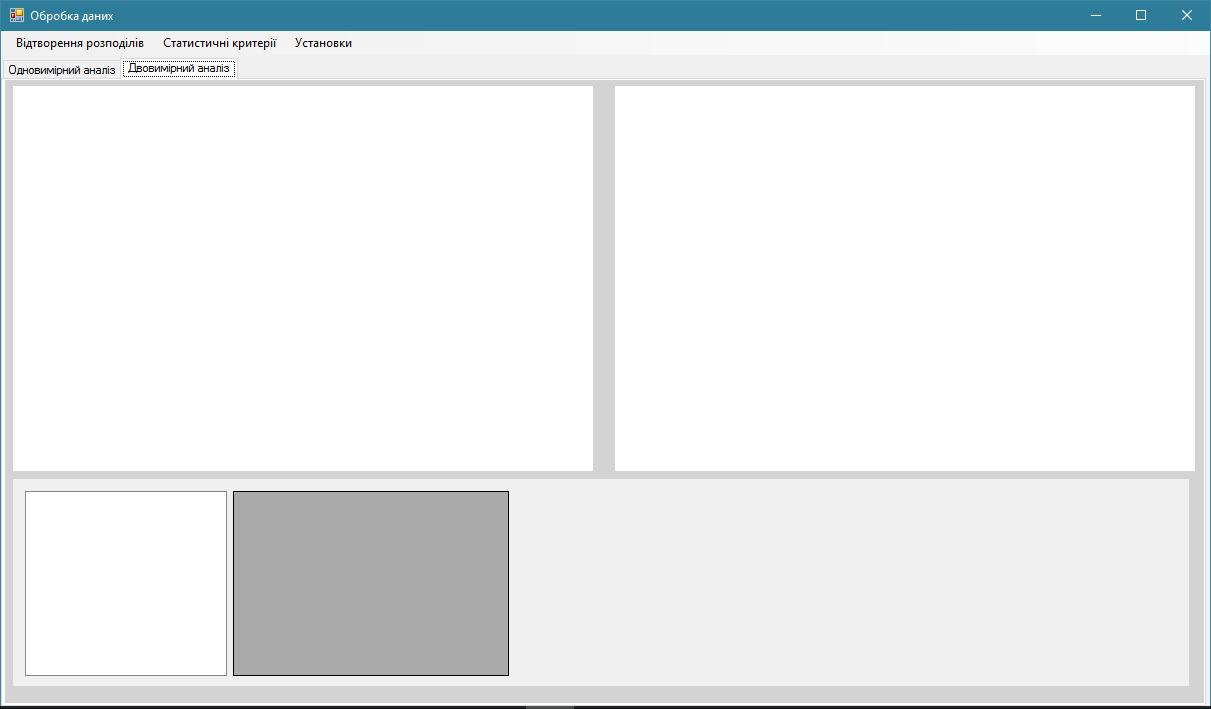




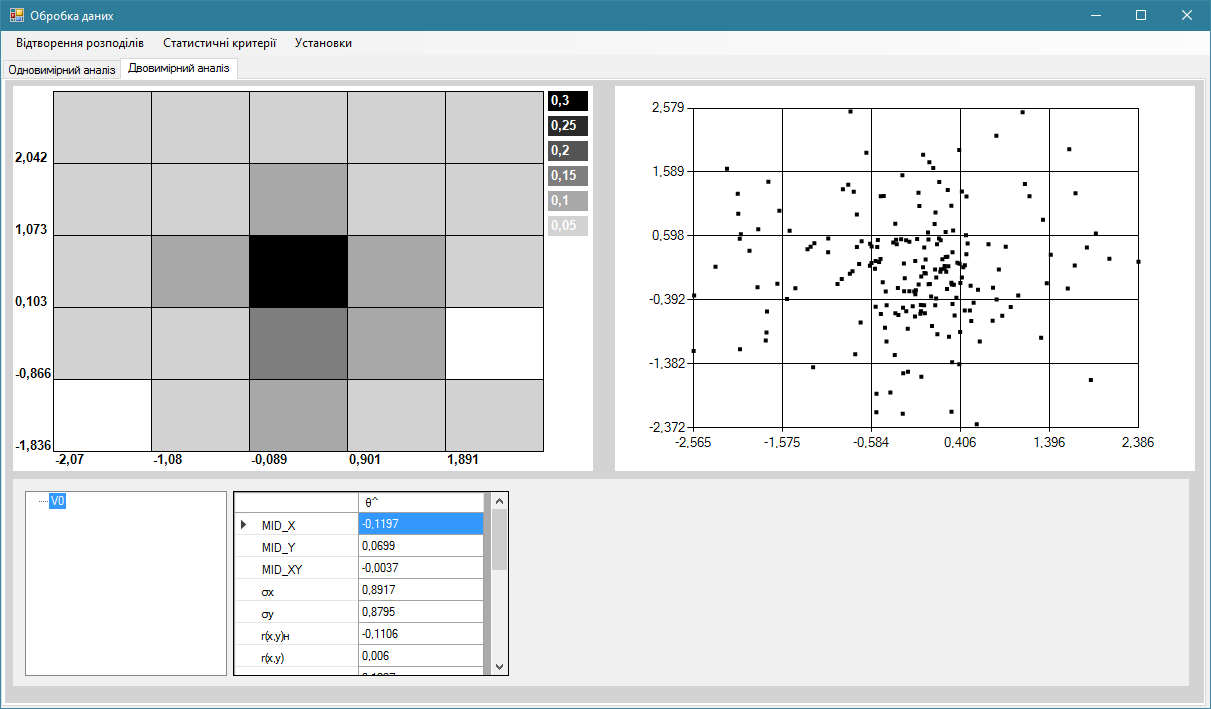




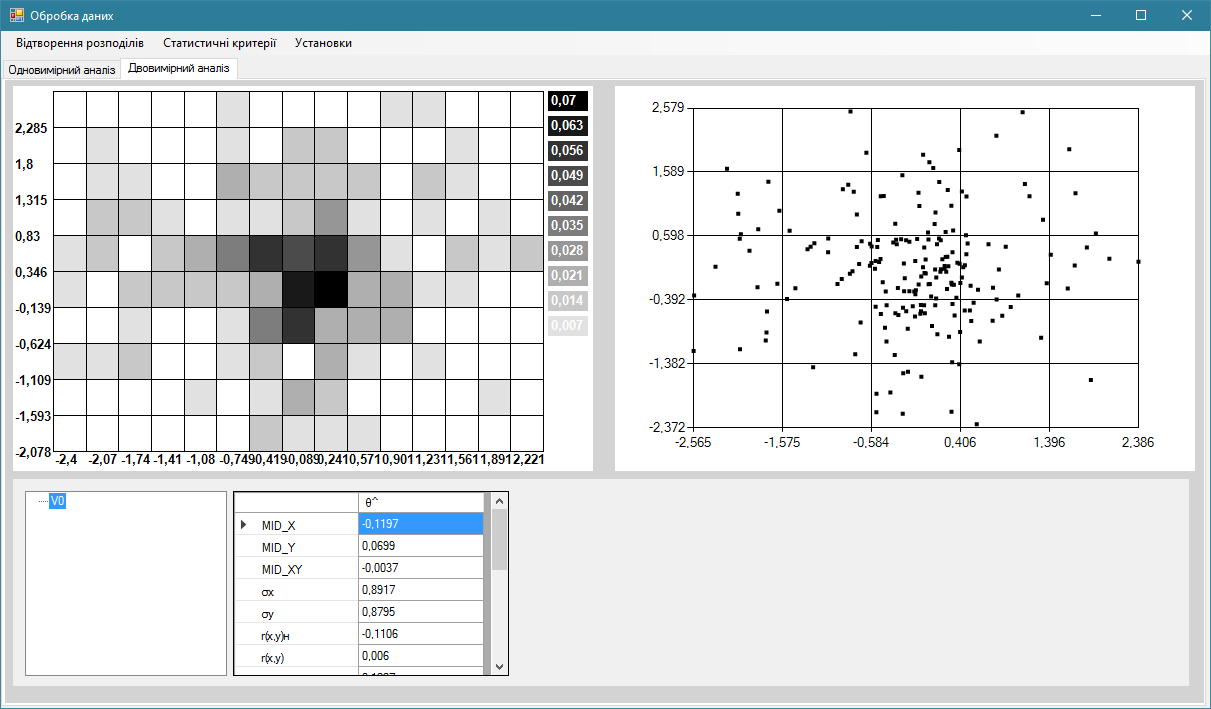
**Практична частина**

Для входу в меню аналізу 2-вимірних даних потрібно обрати пункт меню «2- вимірний аналіз». Обираємо пункт контекстного меню «Згенерувати» чи «Ввести з файлу»

Згенеруємо нормальний 2-вимірний розподіл

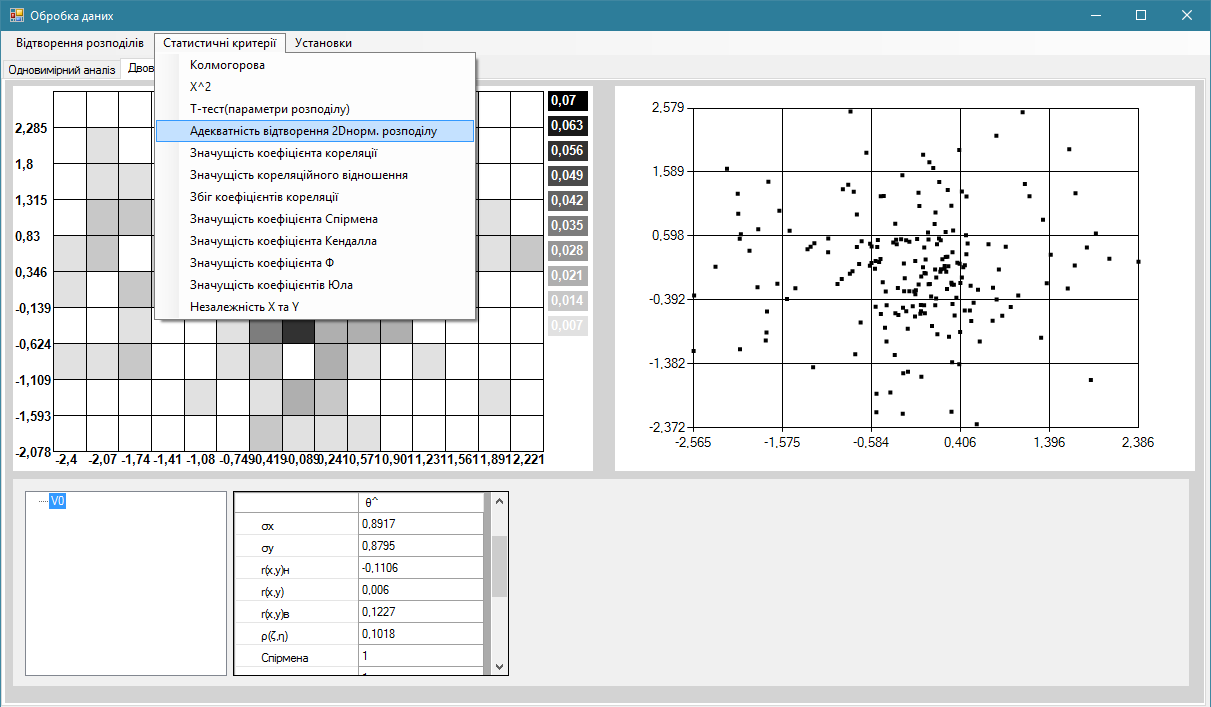


Програма покаже вигляд гістограми та кореляційного поля. Кіліькість класів та класів ймовірності можна задати в установках



Усі коефіцієнти можна та статистики можна переглянути списку оцінок

Щоб перевірити параметри на значущість використаємо меню критеріїв



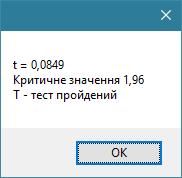


Схема програми

Задання параметрів норм.розподілу

Задання кількості класів

Зчитування з файлу

Генерація вибірки за розподілом

Задання обсягу вибірки

Обрання вибірки

Побудова кореляційного поля та гістограми

Обчислення оцінок

Інтервальні оцінки

Точкові(мат.сподівання, дисперсія)

Кореляційні(коефіцієнт, відношення)

Точкові(мат.сподівання, дисперсія)

Парна рангова кореляція (Спірмана, Кендалла)

Коефіцієнти таблиць 2\*2(Ф, Юла, Фехнера)

Парна рангова кореляція (Спірмана, Кендалла)

Обчислення критеріїв на значущість оцінок

Висновки

1.Я виконав практичну роботу з дисципліни «Аналіз даних». У ній я реалізував можливість роботи з двовимірними даними, генерування двовимірного нормального розподілу, побудову варіаційного ряду.

Після цього програма будує кореляційне поле та вид зверху на двовимірну гістограму.

Також програма обчислює різноманітні оцінки як точкові, кореляційні, парні рангові, сполучень таблиць, та за допомогою критеріїв перевіряє їх на значущість.

При виконанні даної роботи я детально ознайомився з обробкою двовимірних даних, засвоїв знання про коефіцієнт кореляції та кореляційне відношення, коефіцієнти сполучень таблиць та закрепив їх на практиці.

Список використаної літератури

П.О. Приставка, О.М.Мацуга: «АНАЛІЗ ДАНИХ» - Електронний посібник для студентів спеціальності «прикладна математика»