Лабораторна робота

з дисципліни «Аналіз даних» на тему:

«Аналіз багатовимірних даних. Перевірка гіпотез про збіг параметрів багатовимірних даних »

Виконав:

Студент групи ПМ-351

Подлєсний М.Д.

Перевірив:

завідувач кафедри

прикладної математики,

професор

Приставка П.О.

Зміст

[ПОСТАНОВКА ЗАДАЧІ 3](#__RefHeading__713_198435239)

[ТЕОРЕТИЧНА ЧАСТИНА 4](#__RefHeading__715_198435239)

[Первинний статистичний аналіз багатовимірних даних 4](#__RefHeading__717_198435239)

[Перевірка гіпотез про збіг параметрів багатовимірних даних 7](#__RefHeading__2134_198435239)

[ПРАКТИЧНА ЧАСТИНА 11](#__RefHeading__719_198435239)

[Записка користувача 11](#__RefHeading__721_198435239)

[UML-діаграма 13](#__RefHeading__831_198435239)

[ВИСНОВКИ 14](#__RefHeading__725_198435239)

# ПОСТАНОВКА ЗАДАЧІ

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

1. Первинний статистичний аналіз багатовимірних об’єктів спостережень (знаходження векторів середніх та середньоквадратичних, дисперсійно-коваріаційної та кореляційної матриць);
2. Стандартизацію даних (за вимогою користувача);
3. Перевірку збігу векторів середніх та коваріаційних матриць для множини вибірок у припущенні нормального закону розподілу випадкової величини;

# ТЕОРЕТИЧНА ЧАСТИНА

## Первинний статистичний аналіз багатовимірних даних

Головна характеристика *n*-вимірної випадкової величини є функція розподілу ймовірностей :

 (1)

Якщо існує



то функція F(x1,…xn) неперервна праворуч, і при цьому існує функція щільності розподілу ймовірностей



отже



Якщо  незалежні випадкові величини

,

то



або



Із властивості (2) випливає, що процедура знаходження оцінки функції розподілу ймовірностей F(x1,…xn) (оцінки функції щільності F(x1,…xn)) багатовимірної випадкової величини, на разі незалежності одновимірних складових вектора може бути одержана на підставі оцінок маргінальних розподілів:

 (або ), .

Розглянемо забезпечення проведення ймовірнісної оцінки за масивом Ωn,N

реалізацій *n*-вимірної випадкової величини ξ = (ξ1, ξ2, … ξn) з функцією розподілу ймовірностей оцінки відповідних масивів реалізацій

шляхом уведення низки рівномірних розбиттів за осями спостереженняможемо розглядати й можливість ймовірнісної обробки *n*-вимірного варіаційного ряду, розбитого на класи, згідно рівномірного розбиття :



де  - варіанта, яка визначає центральну (або мінімальну) точку -го елементу розбиття ; - кількість елементів (класів) розбиття за напрямками ; - відносна частота варіанти n-вимірного варіаційного ряду:

**Емпірична функція розподілу** Fn,Nза варіаційнимрядом визначається за співвідношенням

**.**

Відмітимо, що для незалежних випадкових величин 



Підготовка даних до аналізу на підставі ймовірнісних оцінок може як містити, так і не містити наступні заходи: стандартизація (нормування) спостережень; при необхідності, зменшення асиметрії в наборах даних; координатні перетворення, що приводять до незалежності окремих складових реалізацій дво- та багатовимірних випадкових величин.

Найпростішими точковими оцінками реалізацій Ωn,N випадкової величини є оцінка вектора математичного сподівання



де



який характеризує геометричний центр тяжіння однорідної сукупності спостережень та оцінка дисперсійно - коваріаційної матриці (ДК-матриці)



де σk2- незсунена оцінка дисперсії *k*-тої ознаки

;

rk,v- оцінка парного коефіціента кореляції поміж *k*-ою та *v*-ою ознаками .

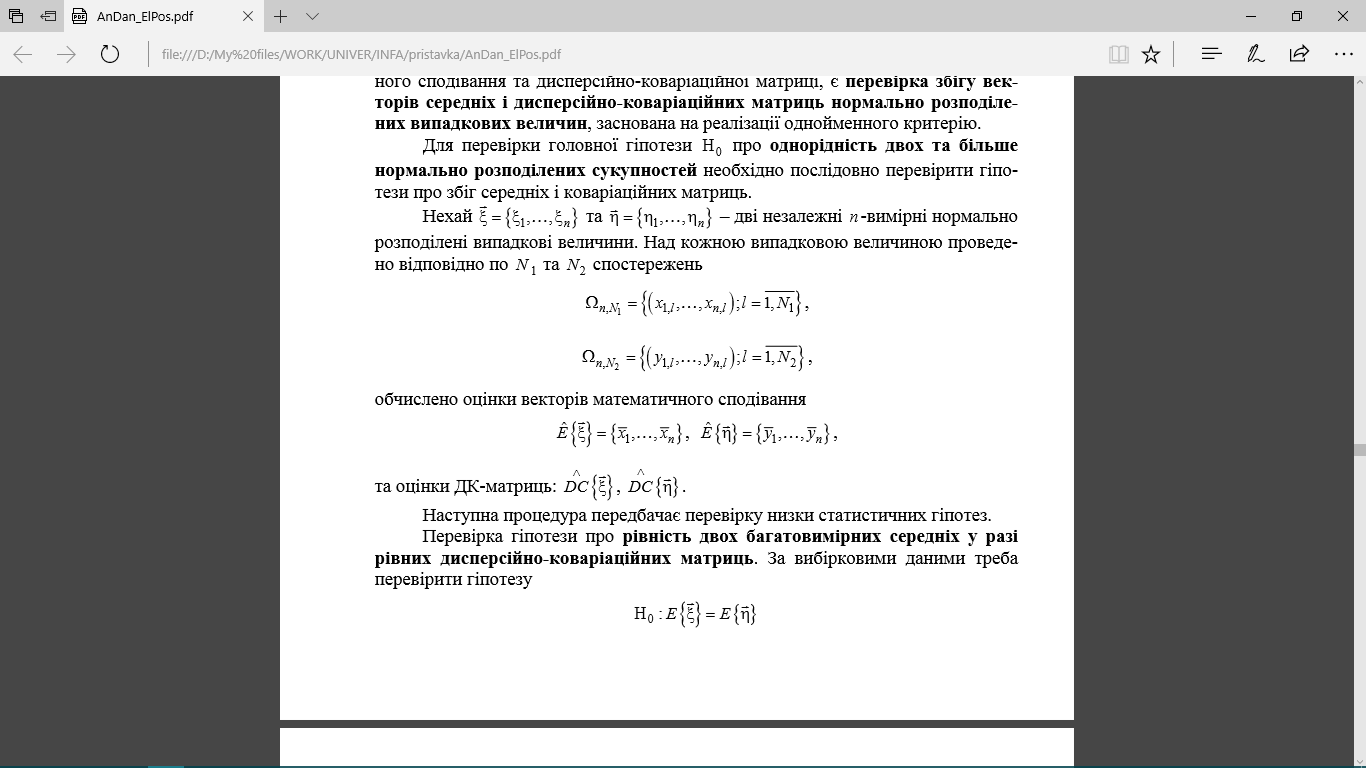
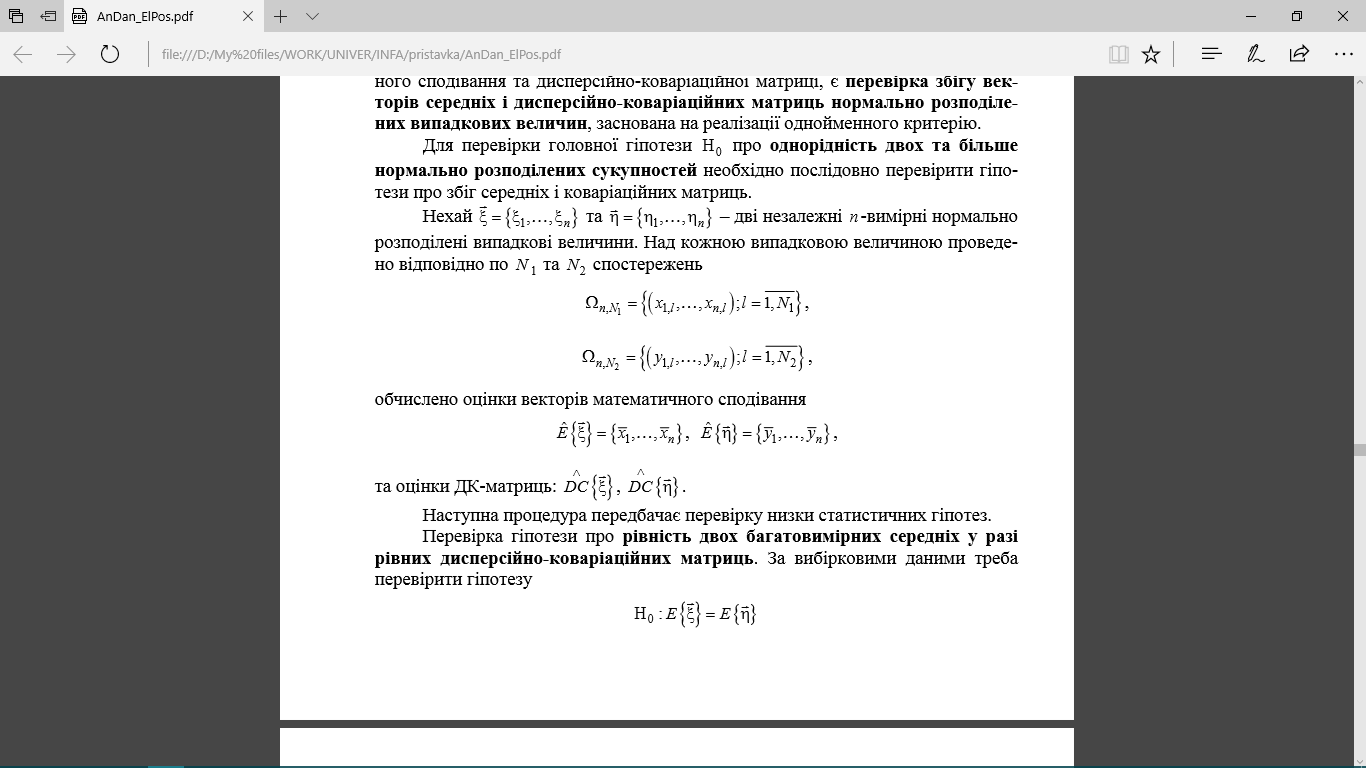
Якщо ознаки об'єкту спостереження розподілені за розподілами, близькими до нормального та вимірюються в різних одиницях масштабу, проводять стандартизацію даних у вигляді

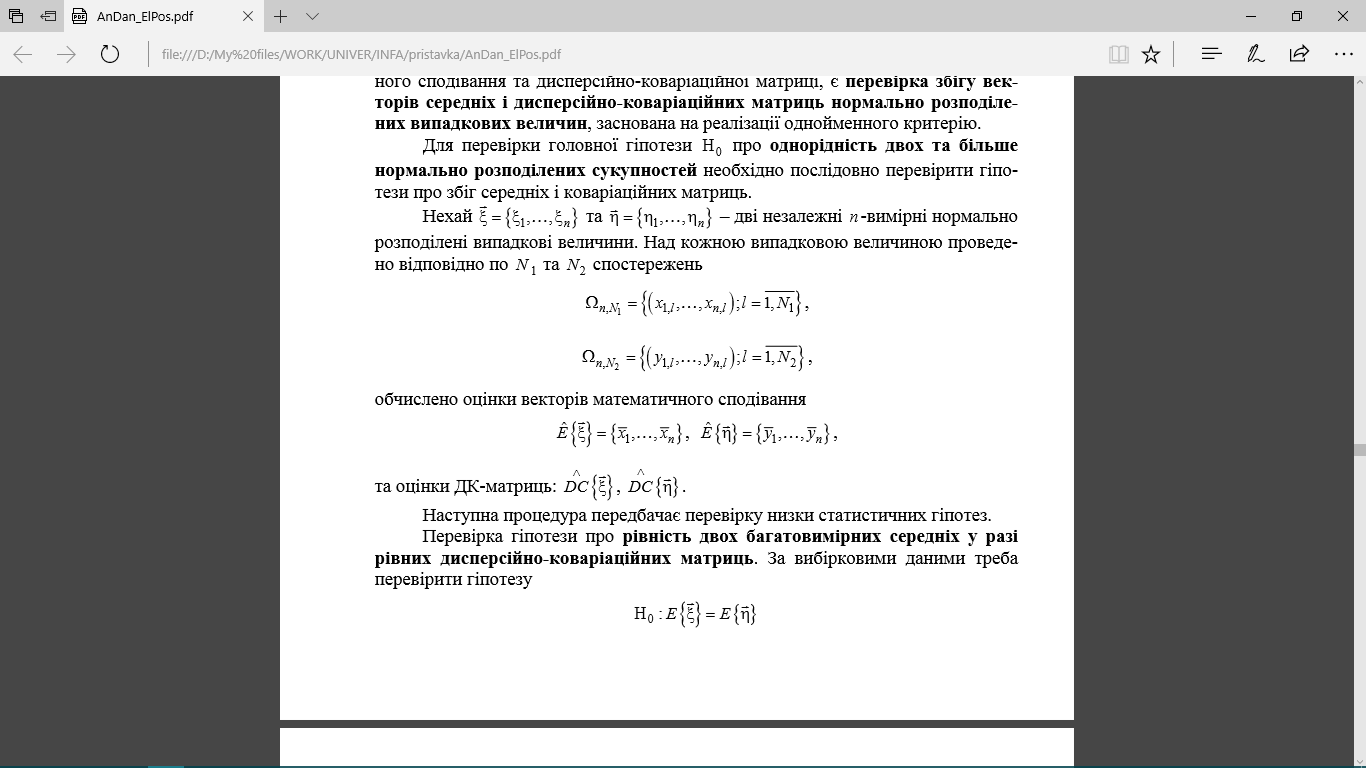


і замість дисперсійно-коваріаційної матриці, оцінці підлягає кореляційна матриця

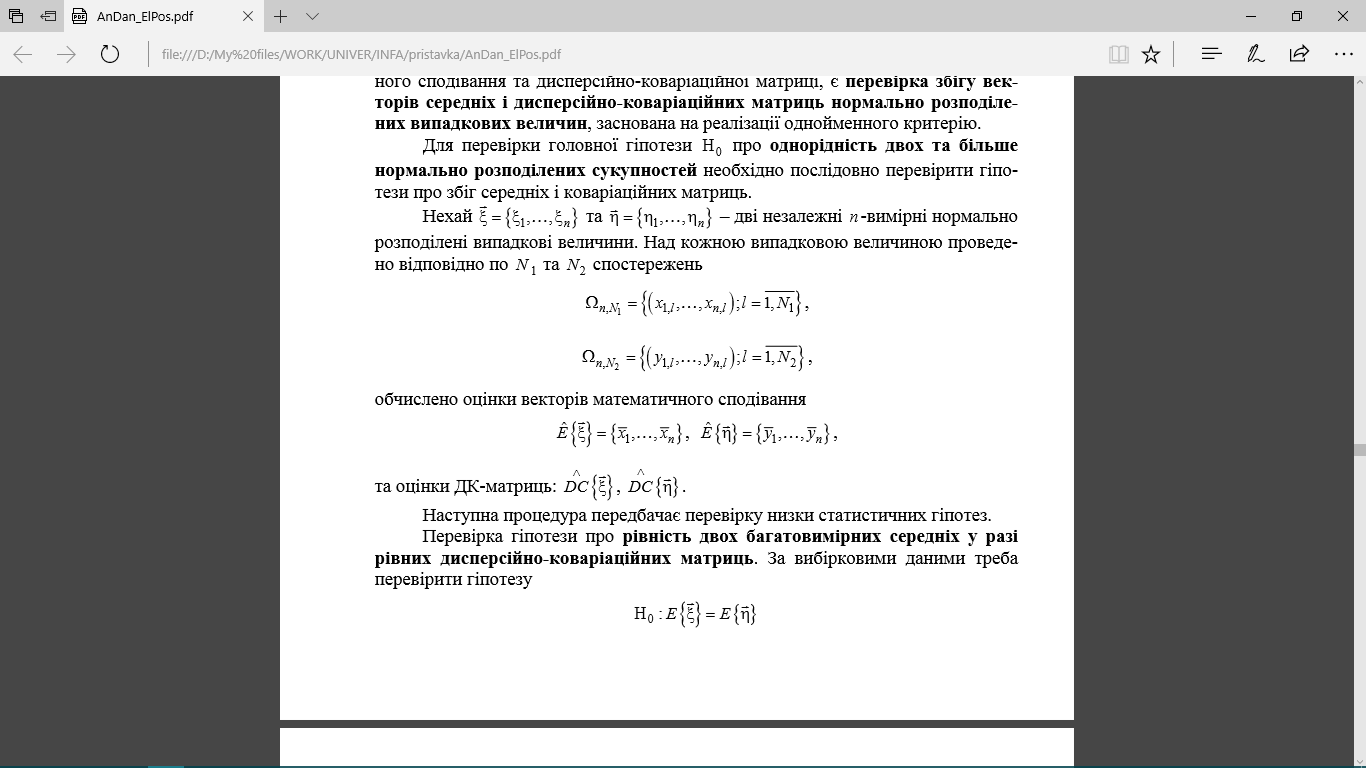


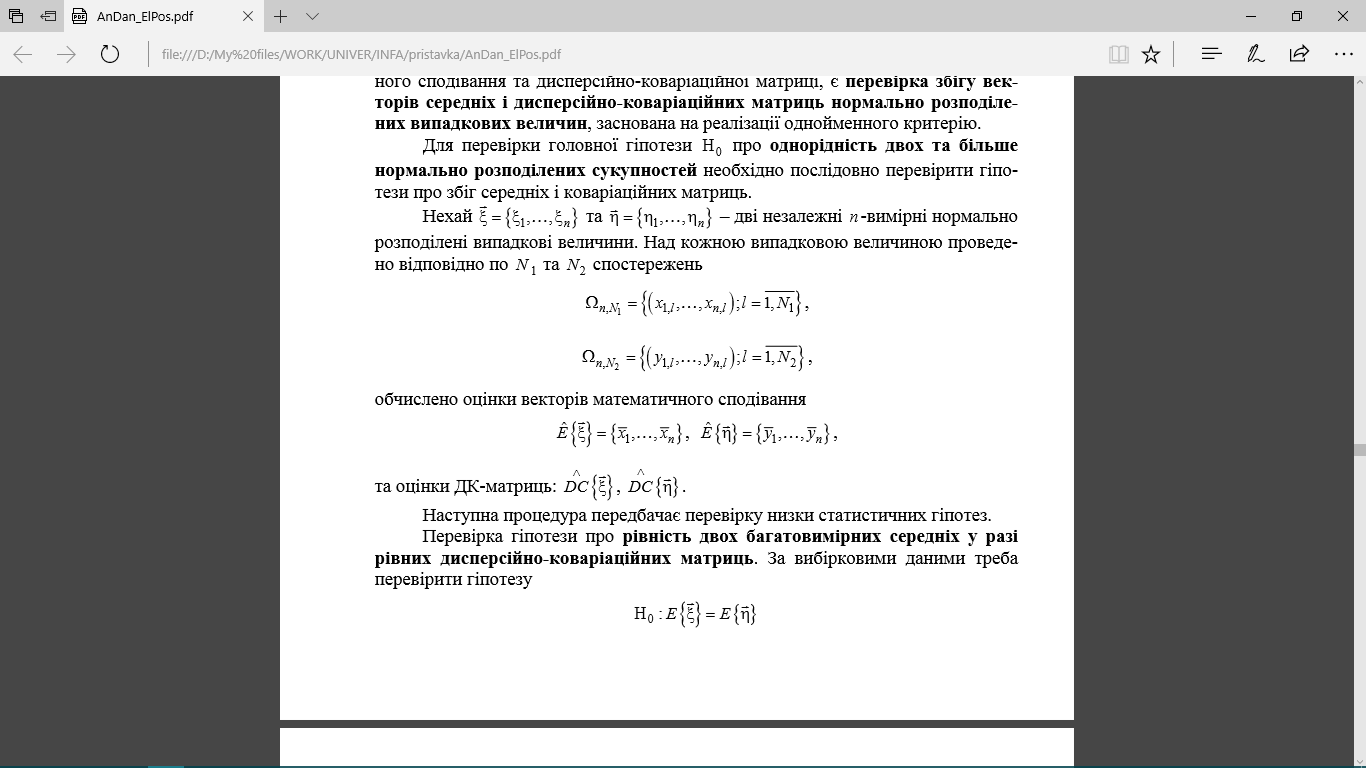
## Перевірка гіпотез про збіг параметрів багатовимірних даних

Нехай та  – дві незалежні n-вимірні нормально розподілені випадкові величини. Над кожною випадковою величиною проведено відповідно по N1 та N2

,

обчислено оцінки векторів математичного сподівання

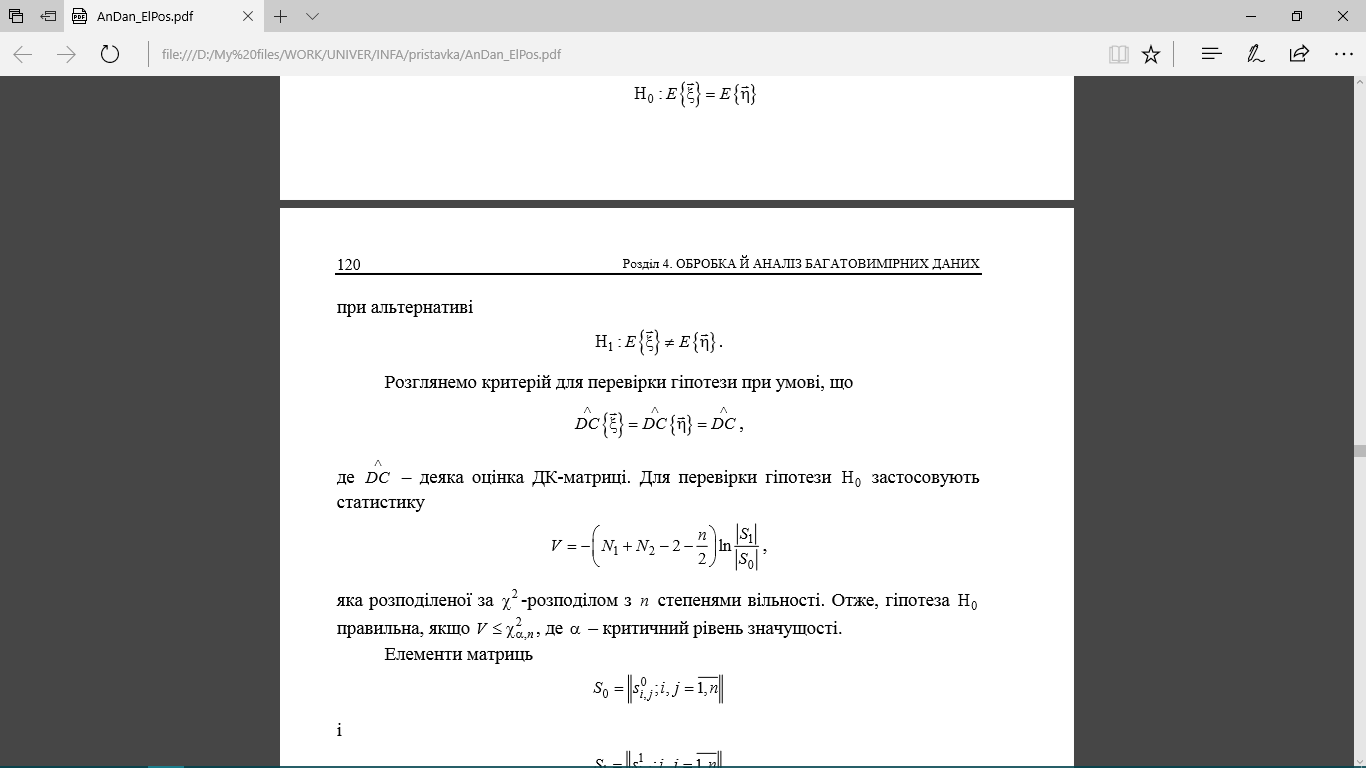
,

та оцінки ДК-матриць: 

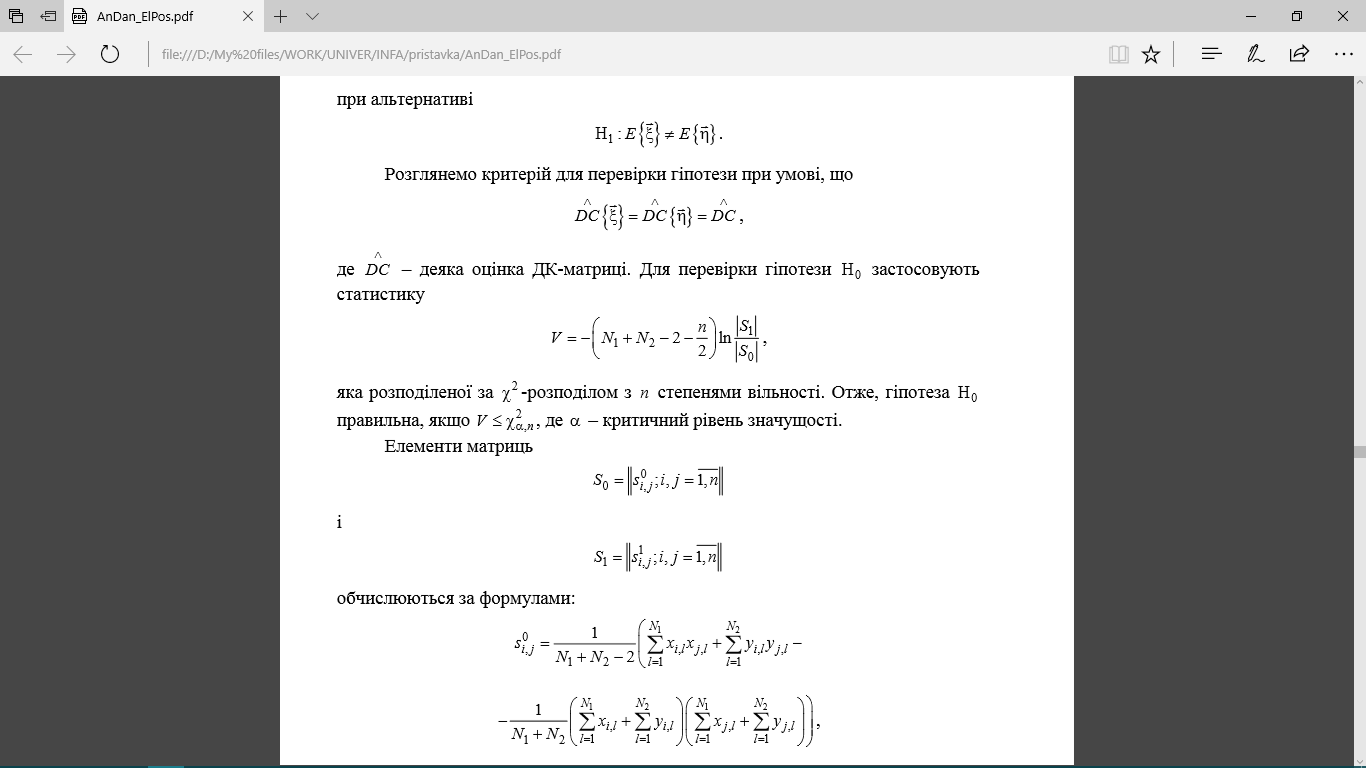
Наступна процедура передбачає перевірку низки статистичних гіпотез.

**Перевірка гіпотези про рівність двох багатовимірних середніх у разі рівних дисперсійно-коваріаційних матриць**.

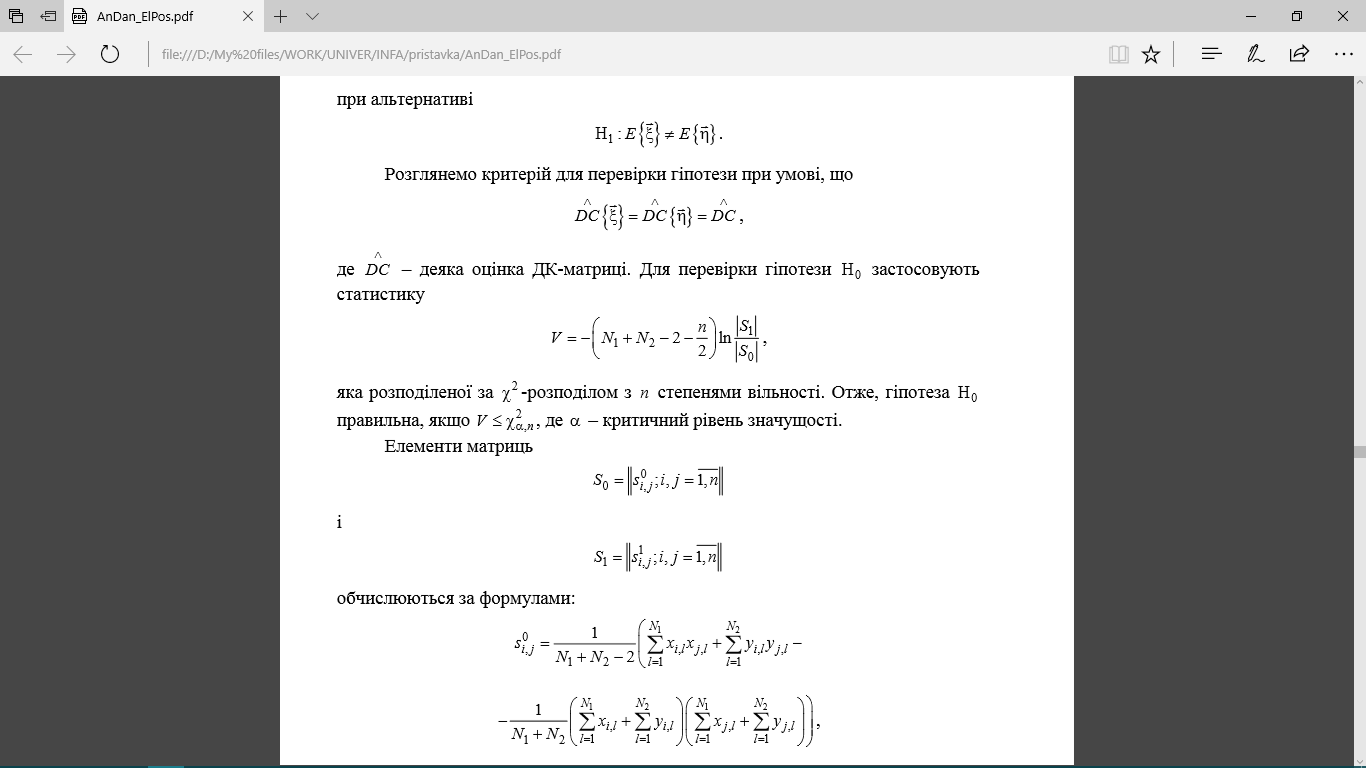
За вибірковими даними треба перевірити гіпотезу

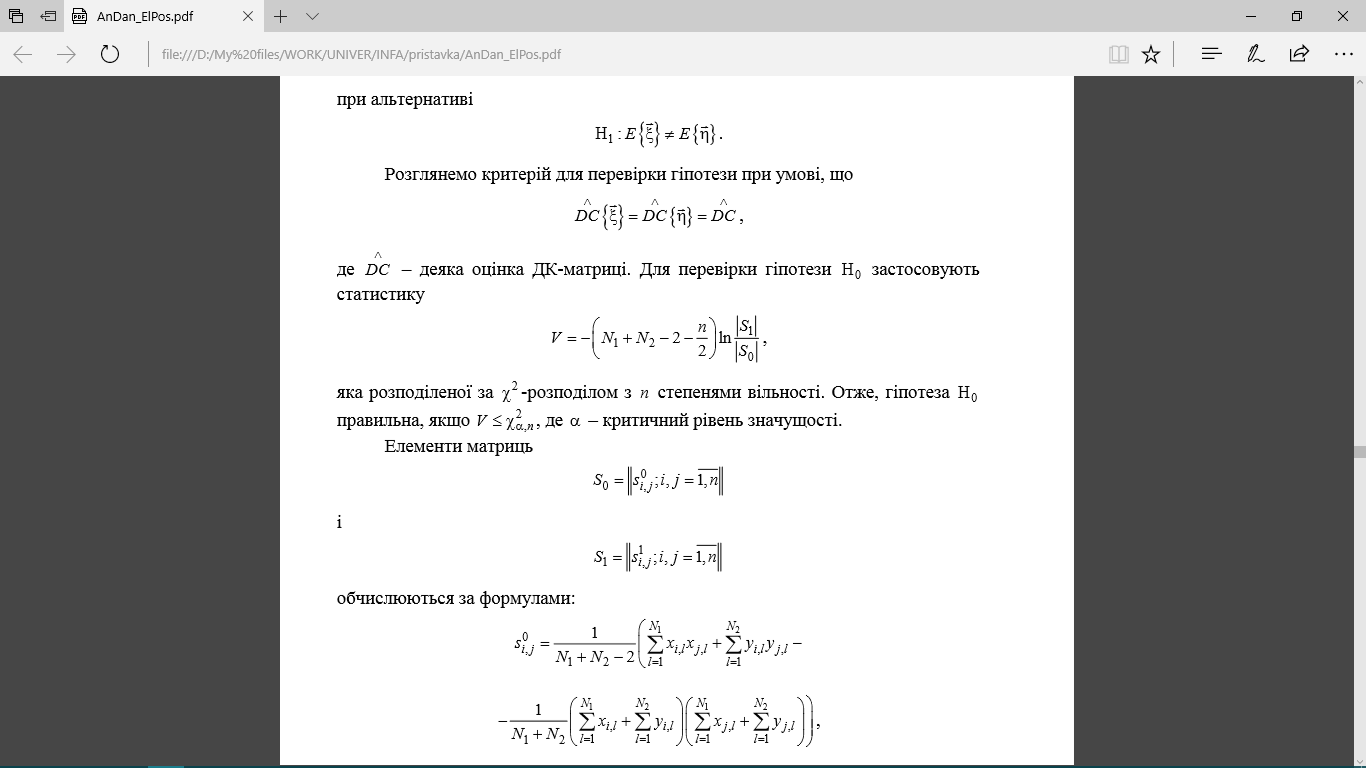


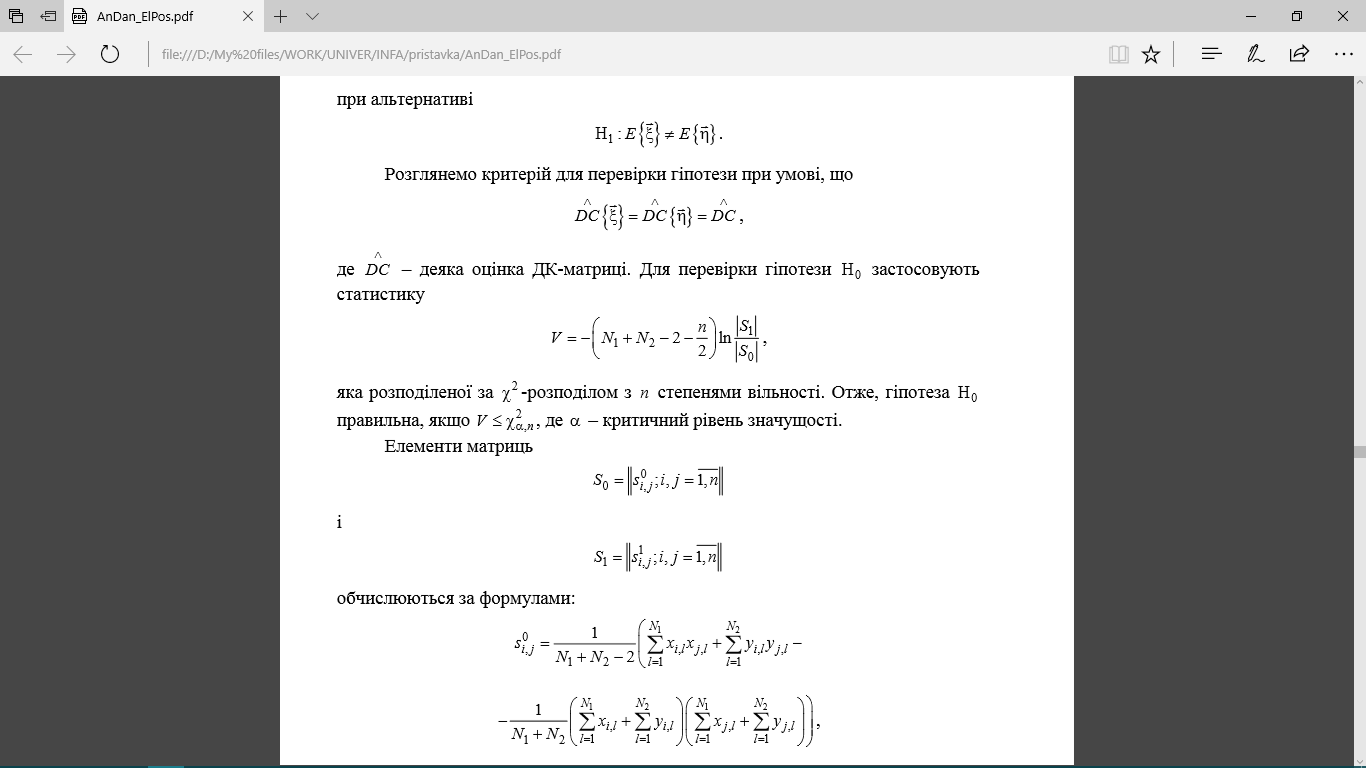
при альтернативі

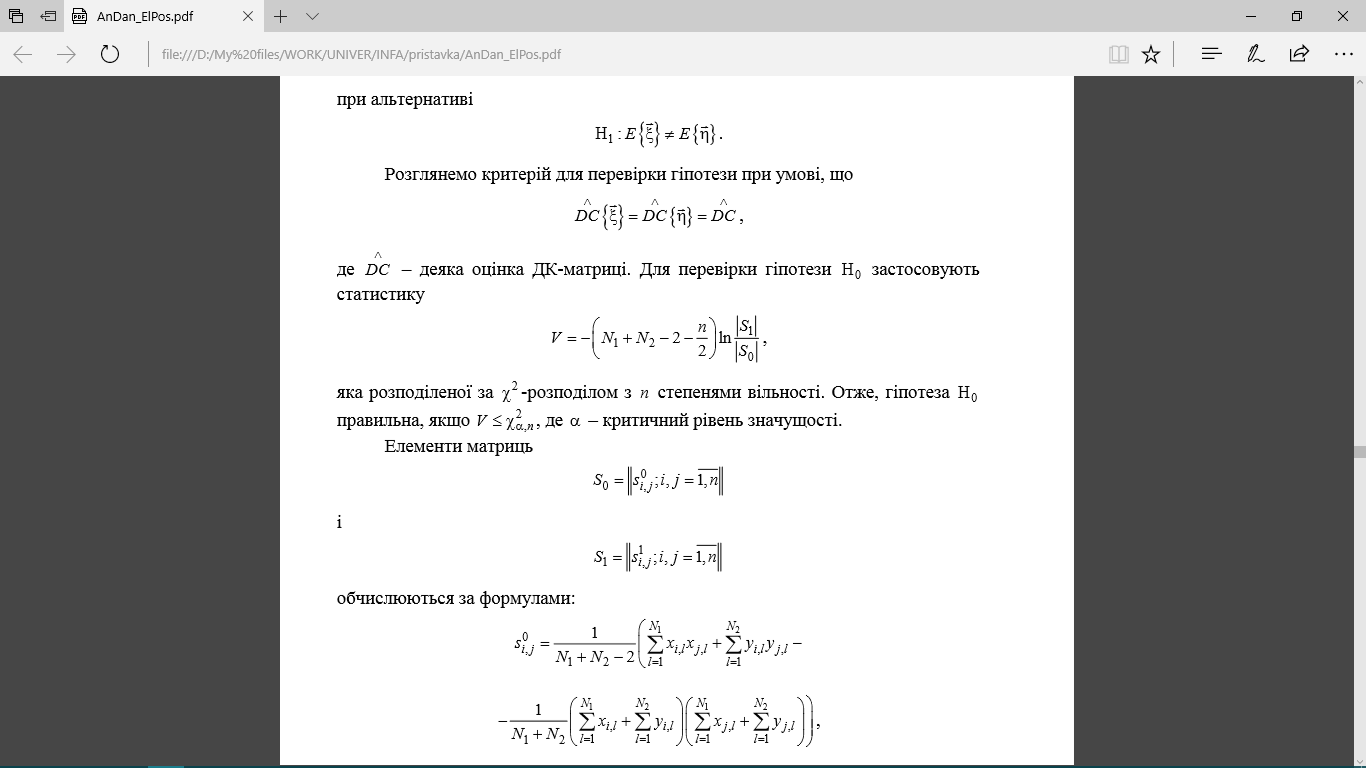
.

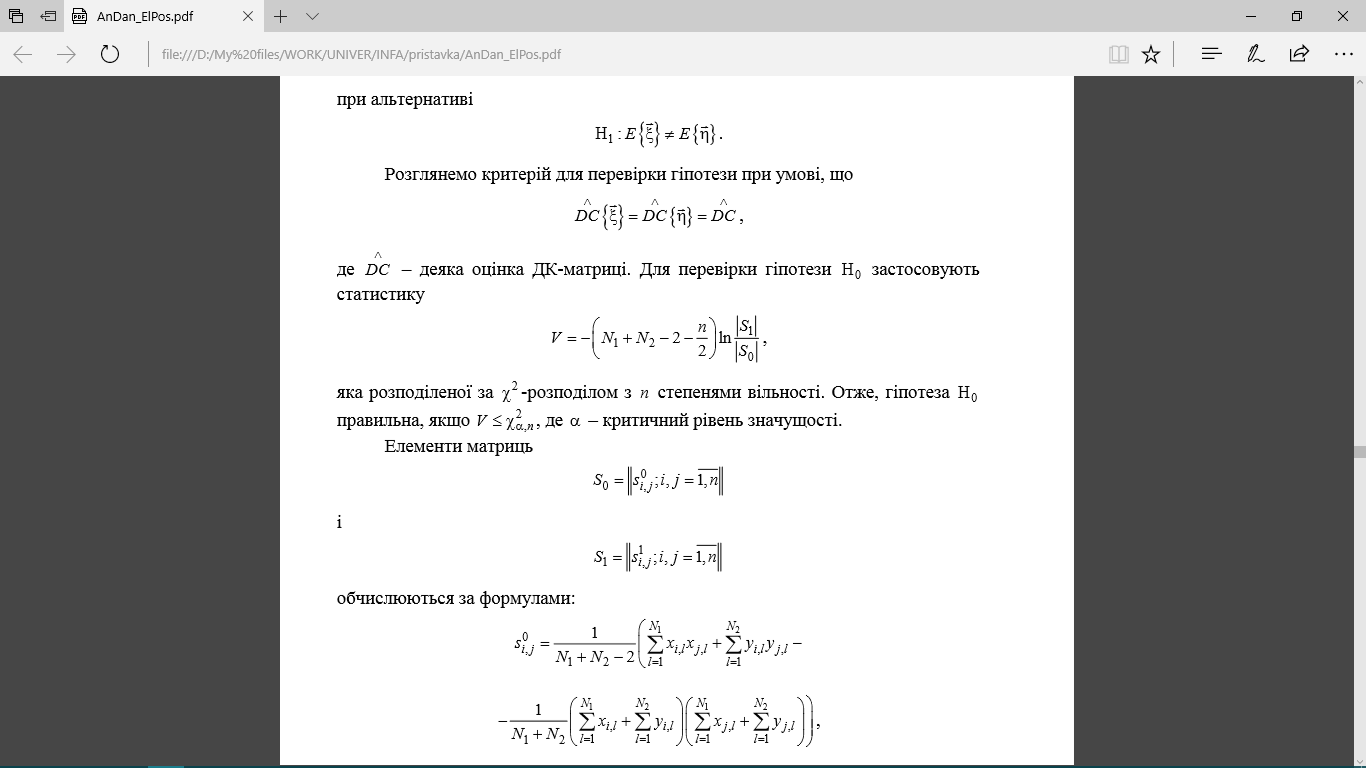
Розглянемо критерій для перевірки гіпотези при умові, що



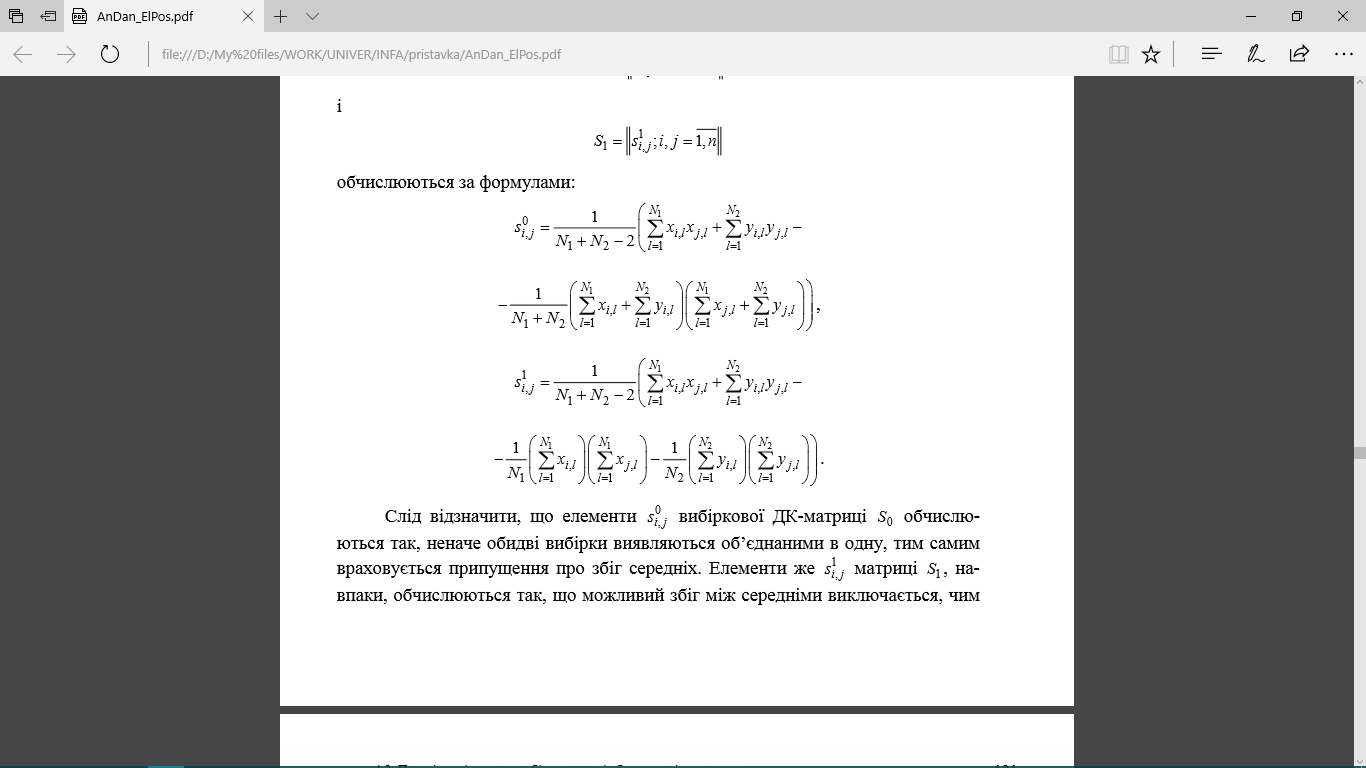
де  – деяка оцінка ДК-матриці. Для перевірки гіпотези Η0 застосовують статистику

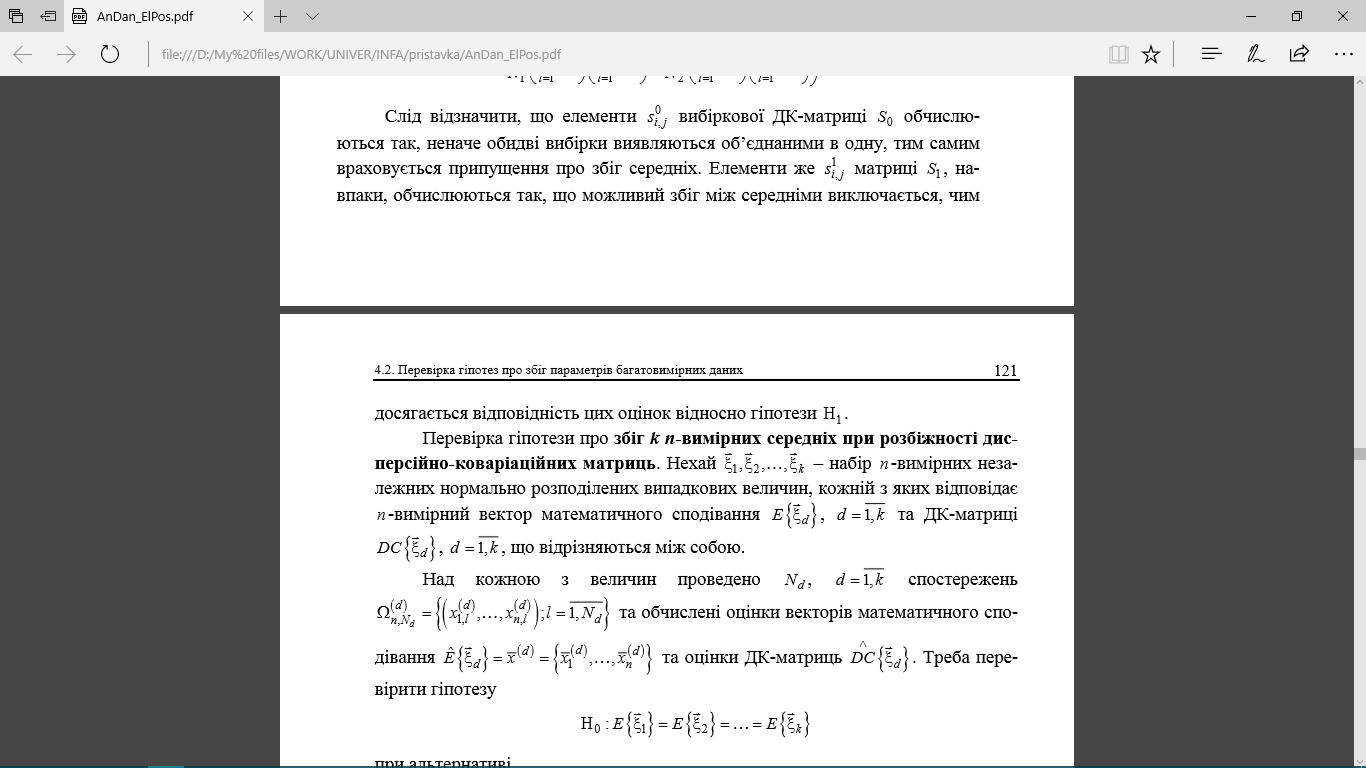
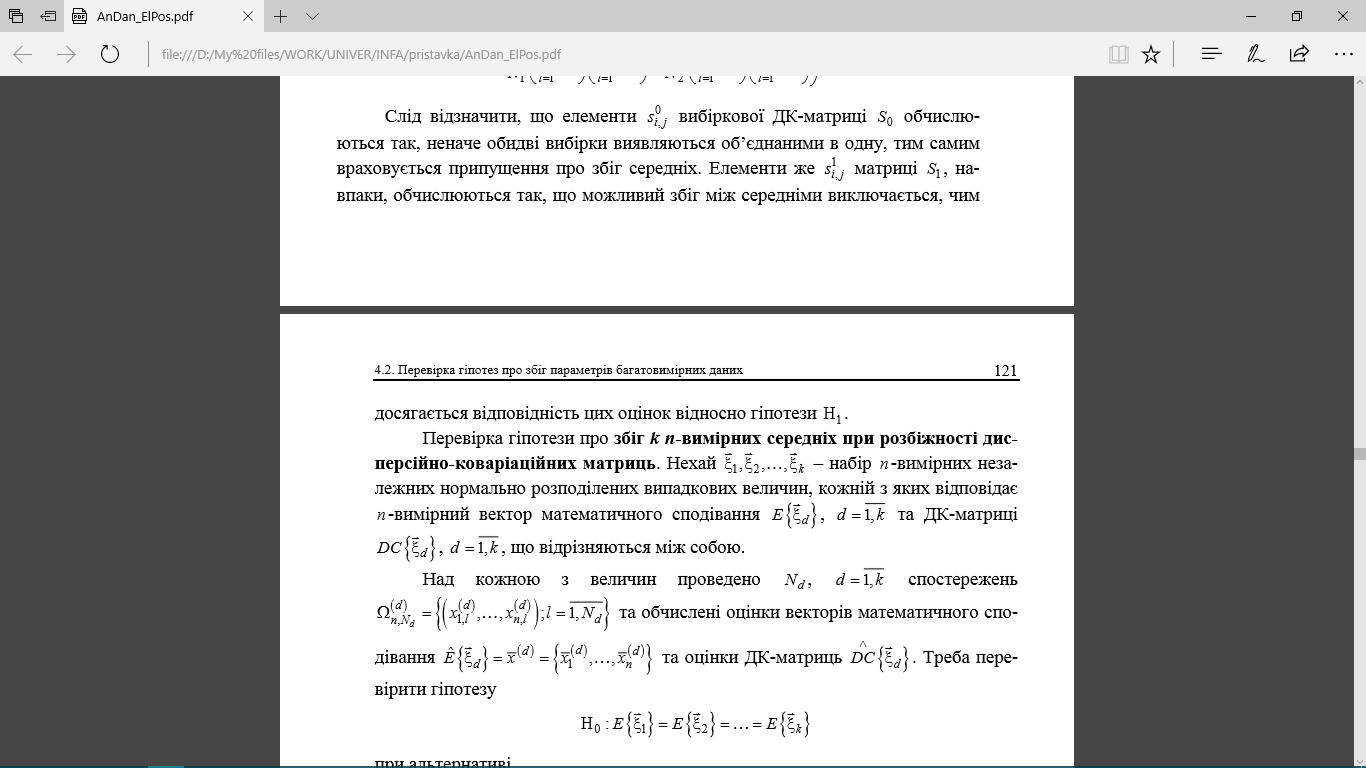


яка розподіленої за χ2 -розподілом з n степенями вільності. Отже, гіпотеза Η0 правильна, якщо  , де α – критичний рівень значущості. Елементи

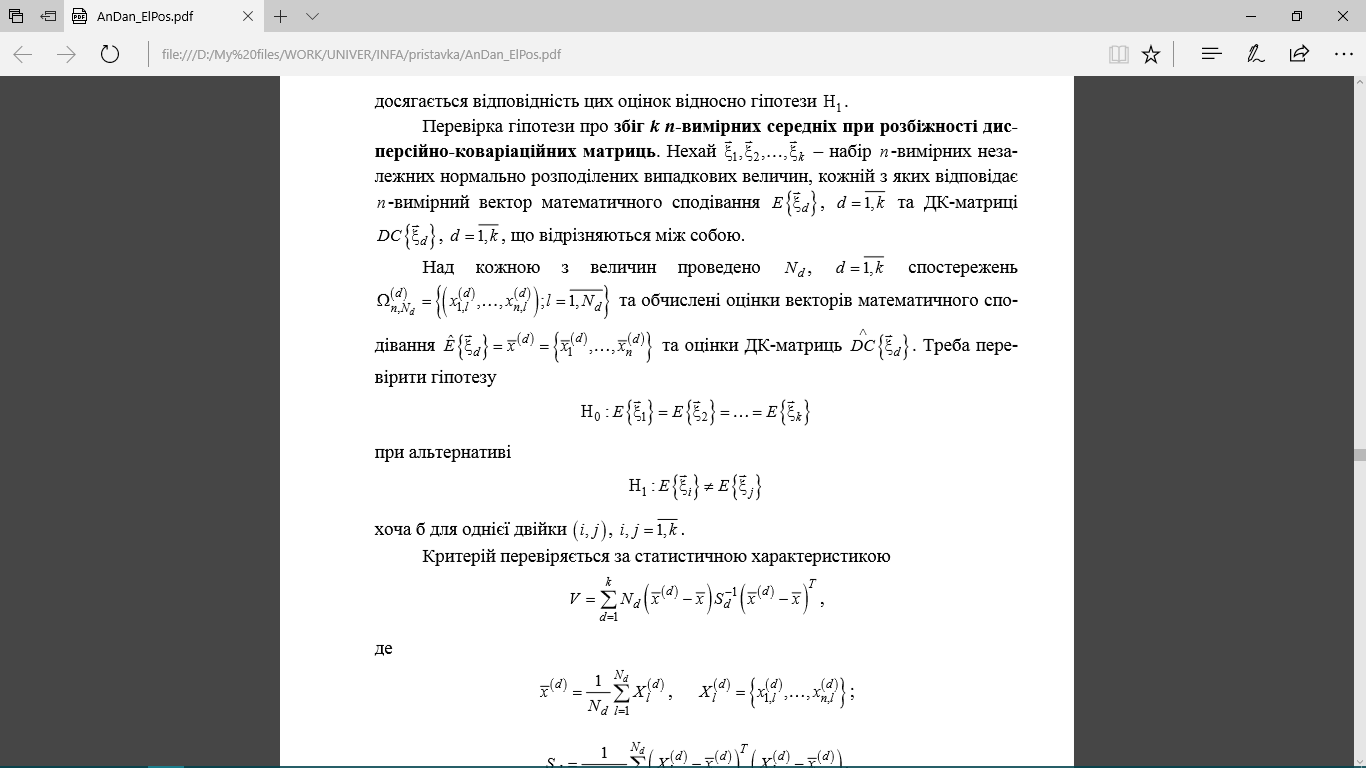
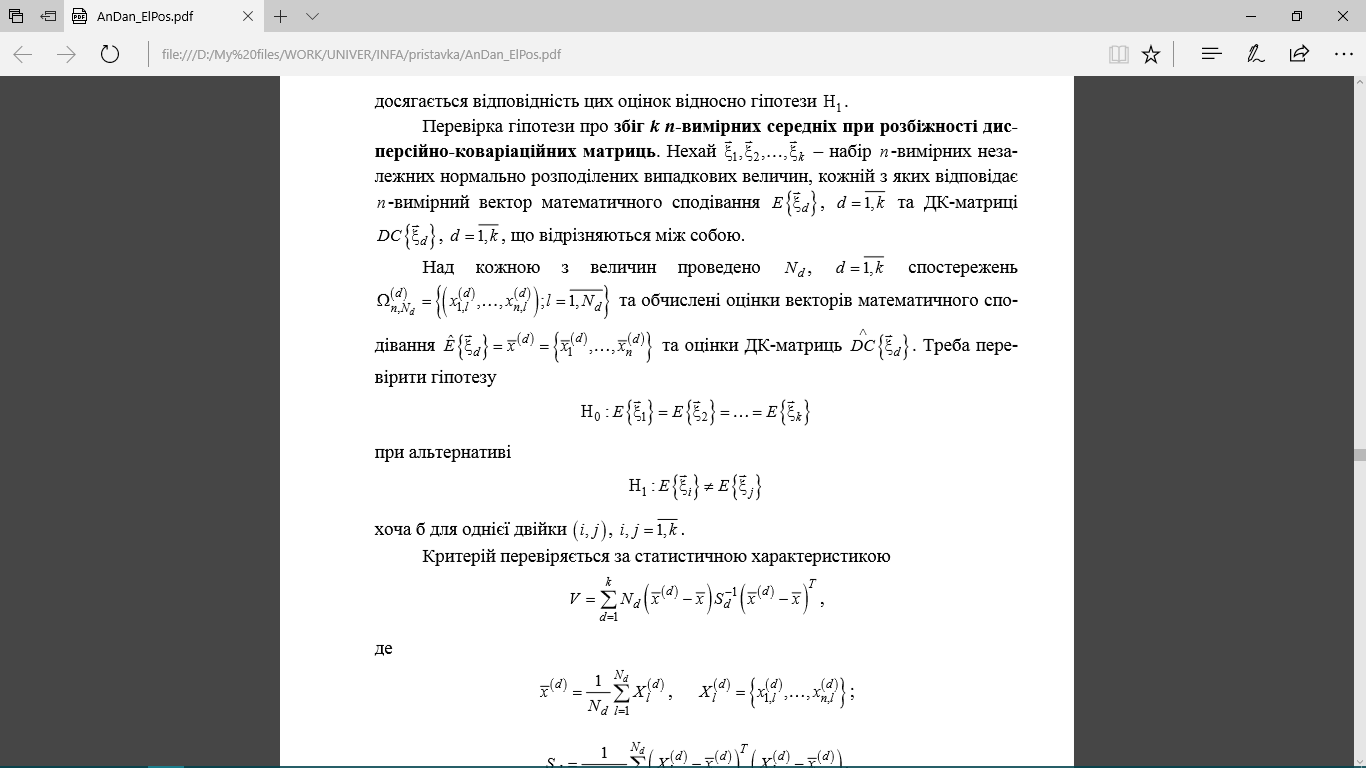
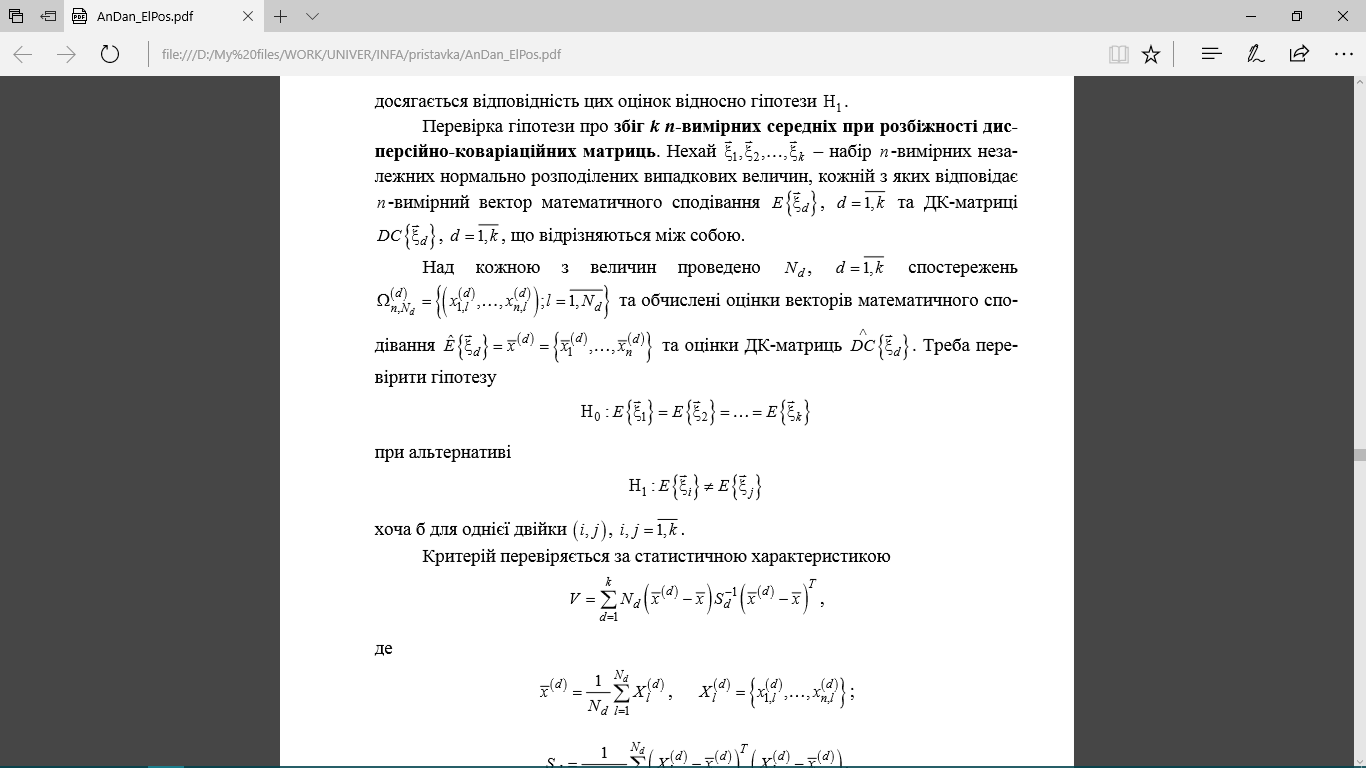
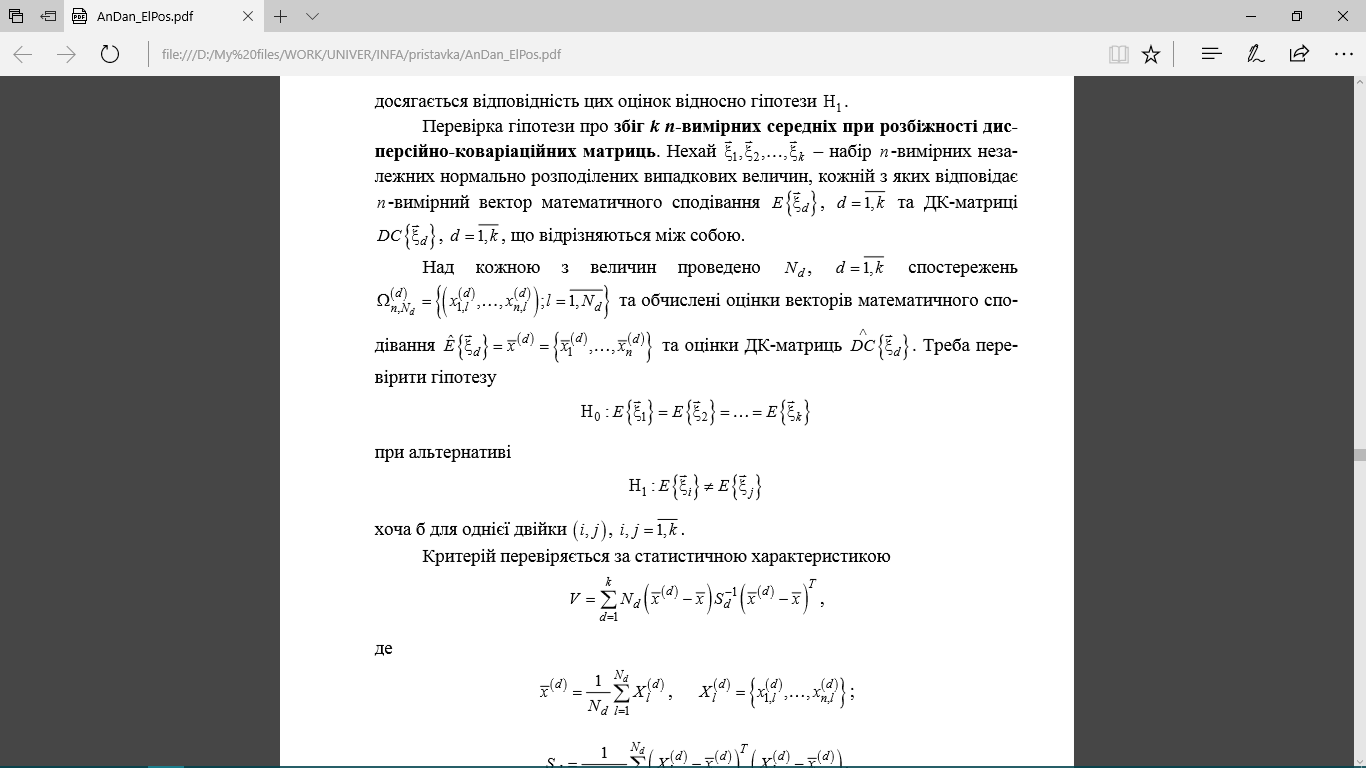
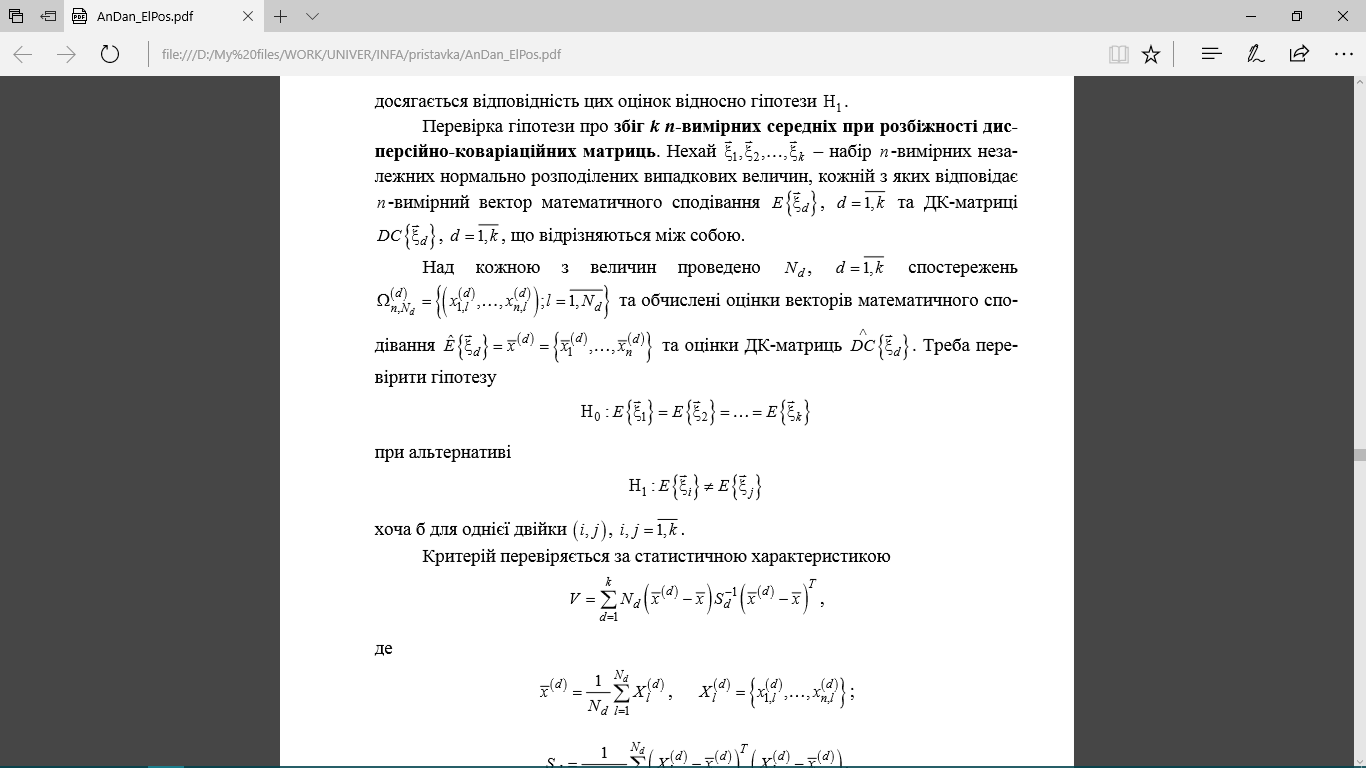
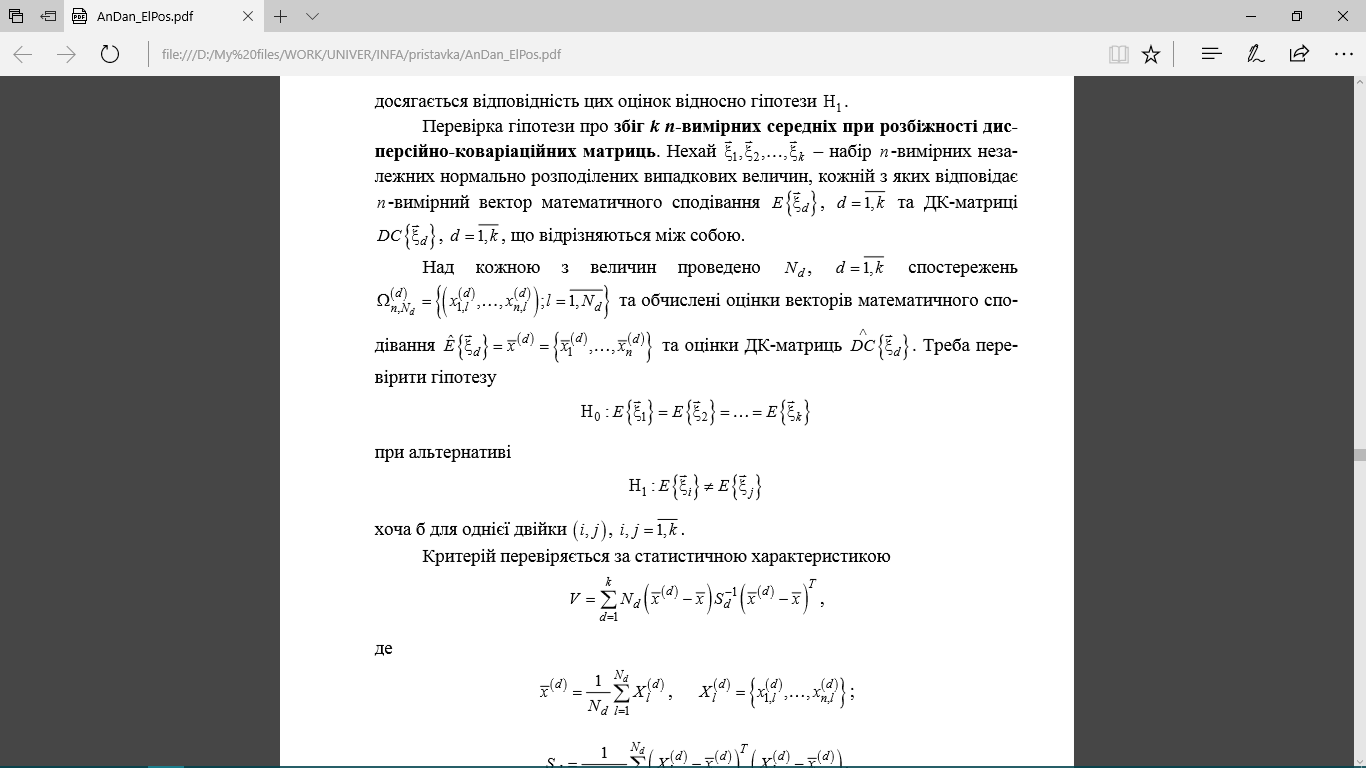


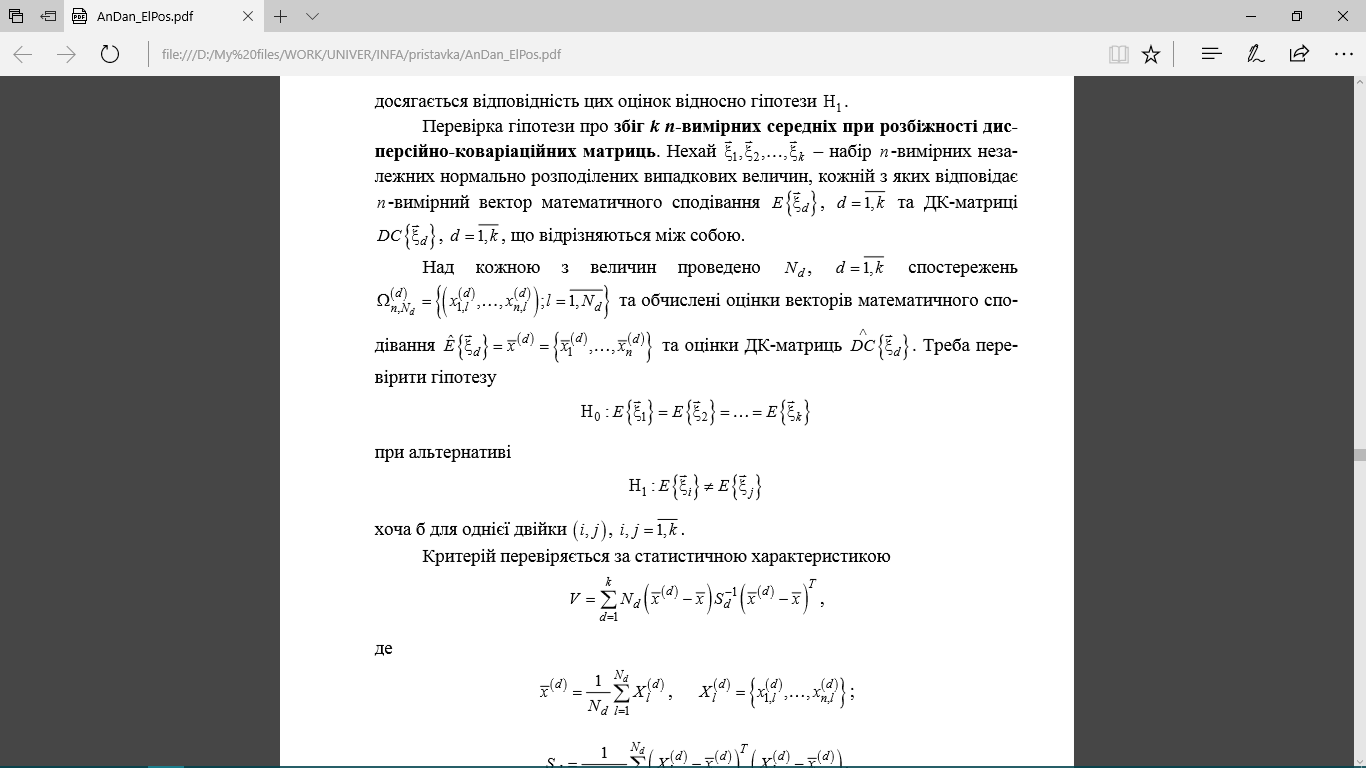
обчислюються за формулами



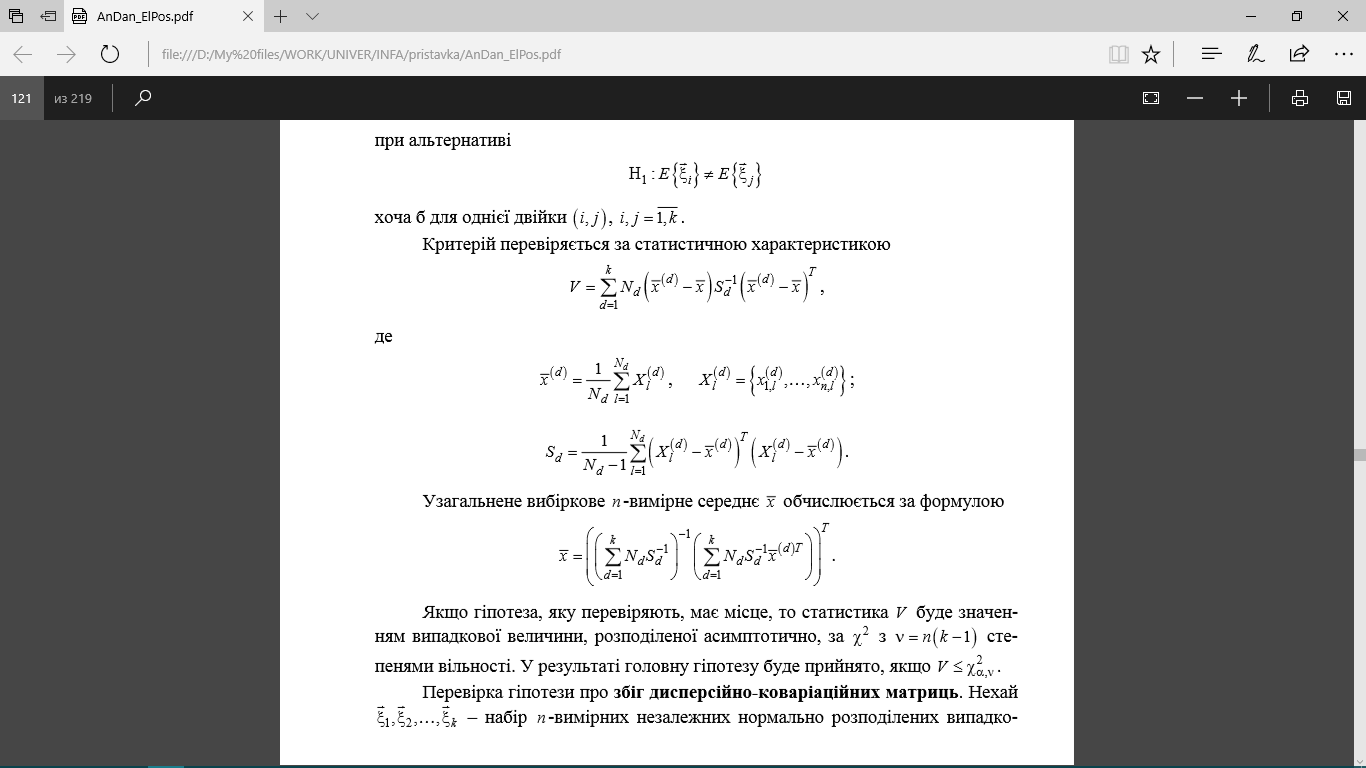
Слід відзначити, що елементи вибіркової ДК-матриці S0 обчислюються так, неначе обидві вибірки виявляються об’єднаними в одну, тим самим враховується припущення про збіг середніх. Елементи же  матриці S1 , навпаки, обчислюються так, що можливий збіг між середніми виключається, чим досягається відповідність цих оцінок відносно гіпотези Η1 .

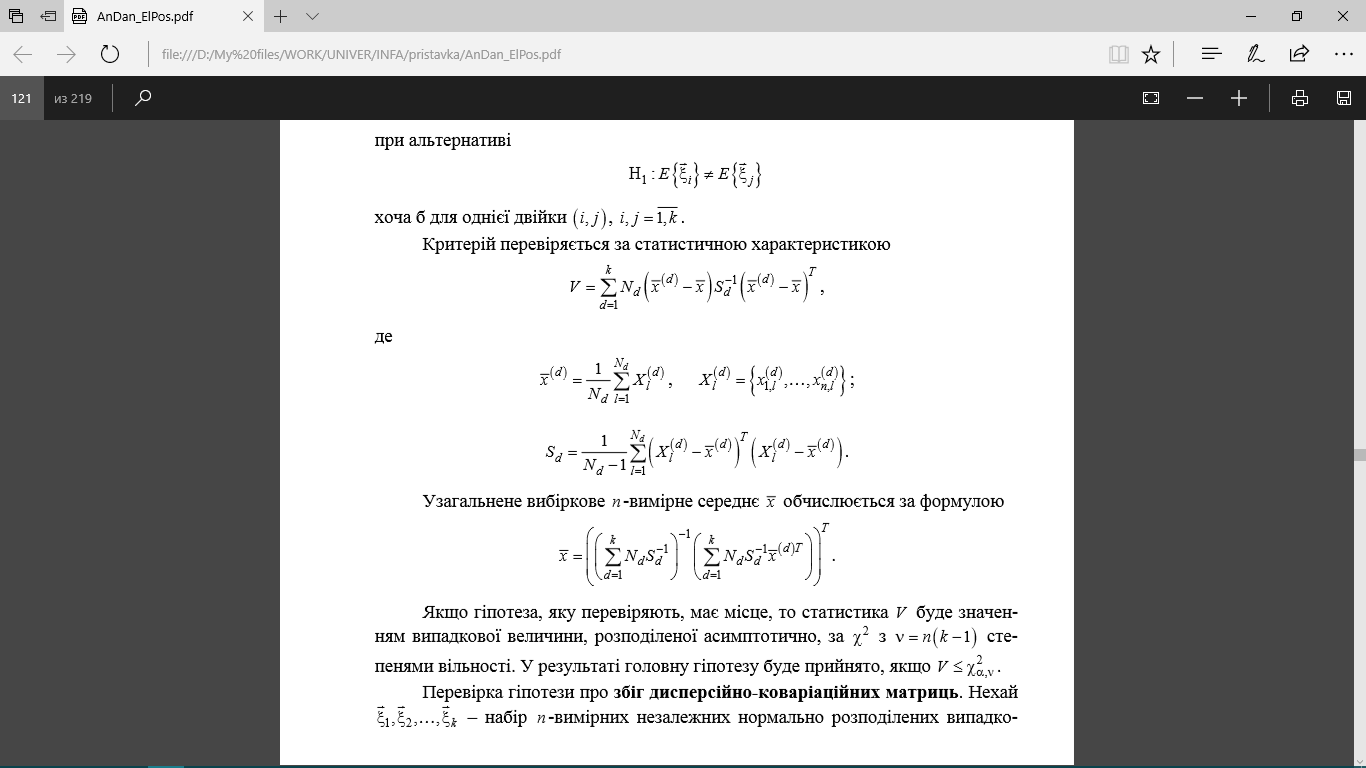
**Перевірка гіпотези про збіг k n-вимірних середніх при розбіжності дисперсійно-коваріаційних матриць**.

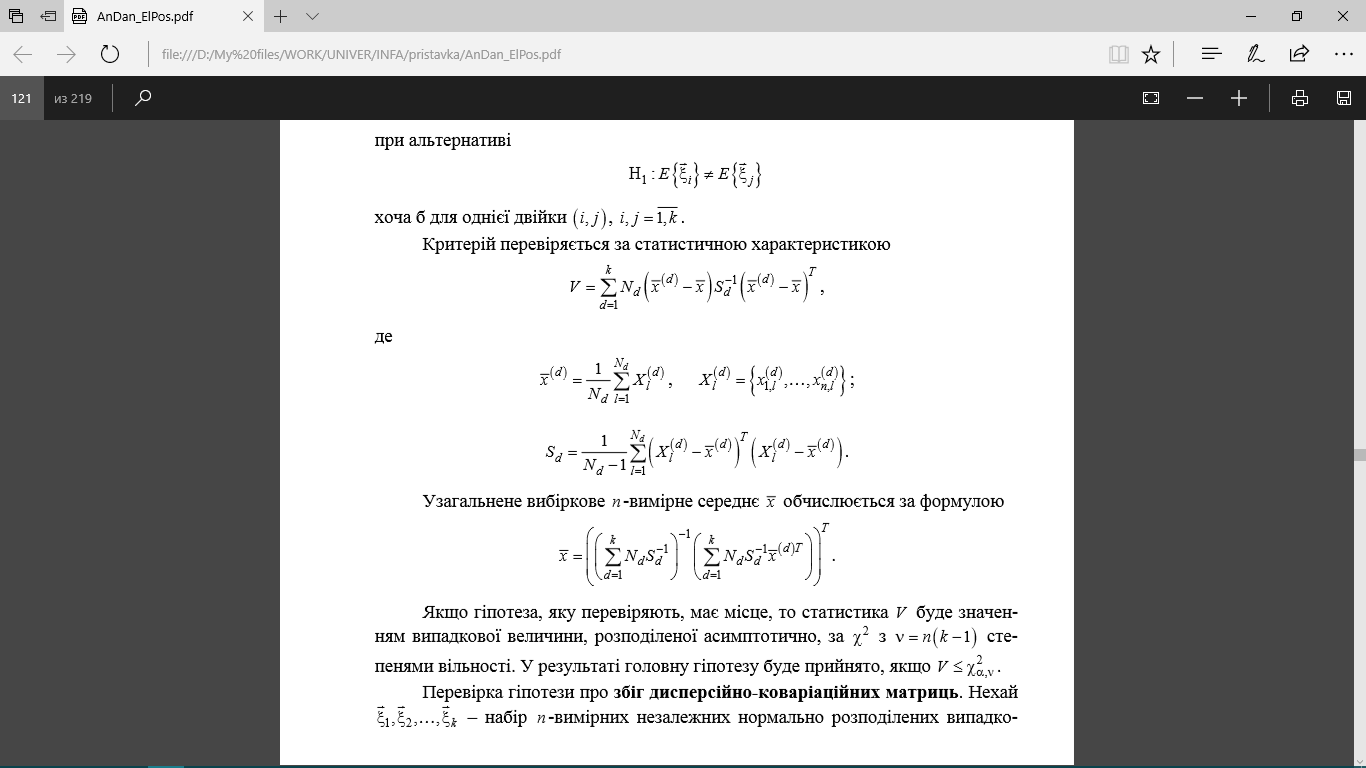
Нехай – набір n-вимірних незалежних нормально розподілених випадкових величин, кожній з яких відповідає n-вимірний вектор математичного сподівання та ДК-матриці, що відрізняються між собою. Над кожною з величин проведено спостережень та обчислені оцінки векторів математичного сподівання та оцінки ДК-матриць . Треба перевірити гіпотезу



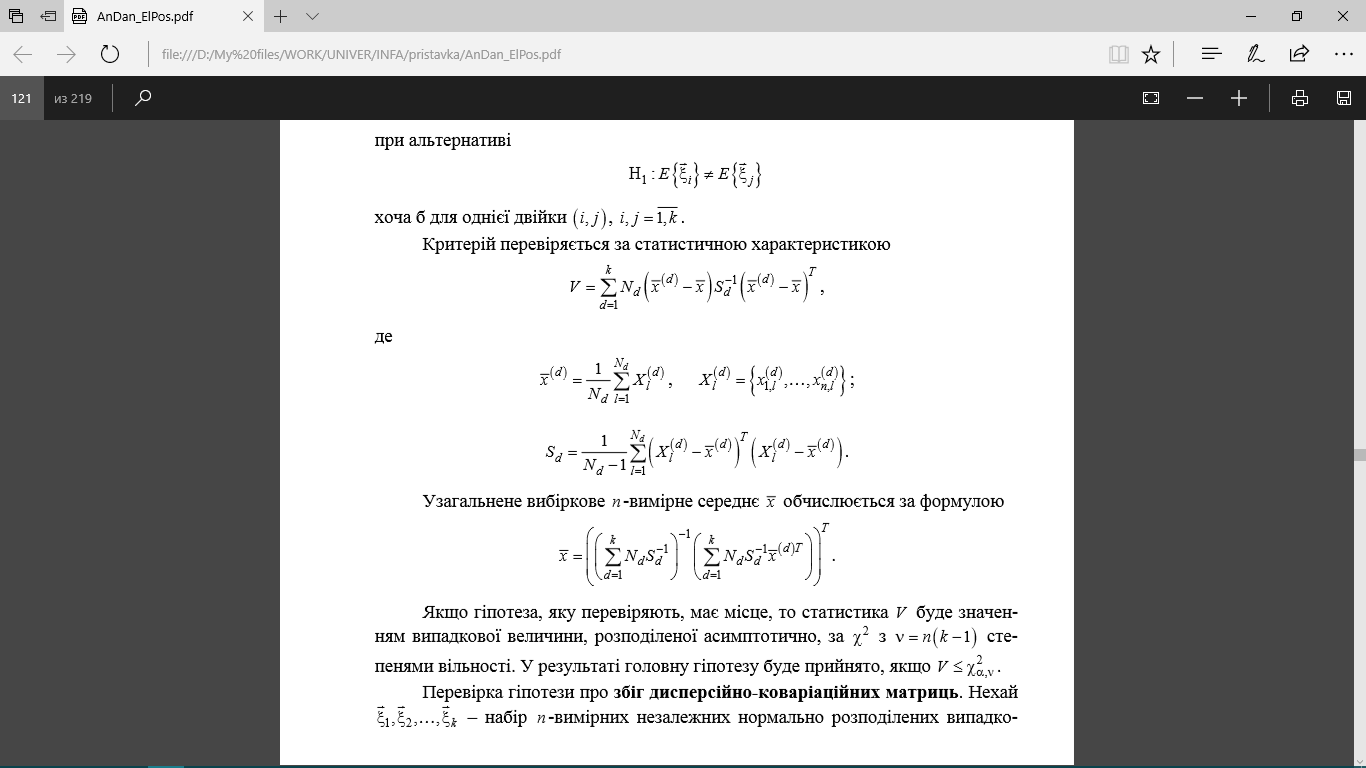
при альтернативі



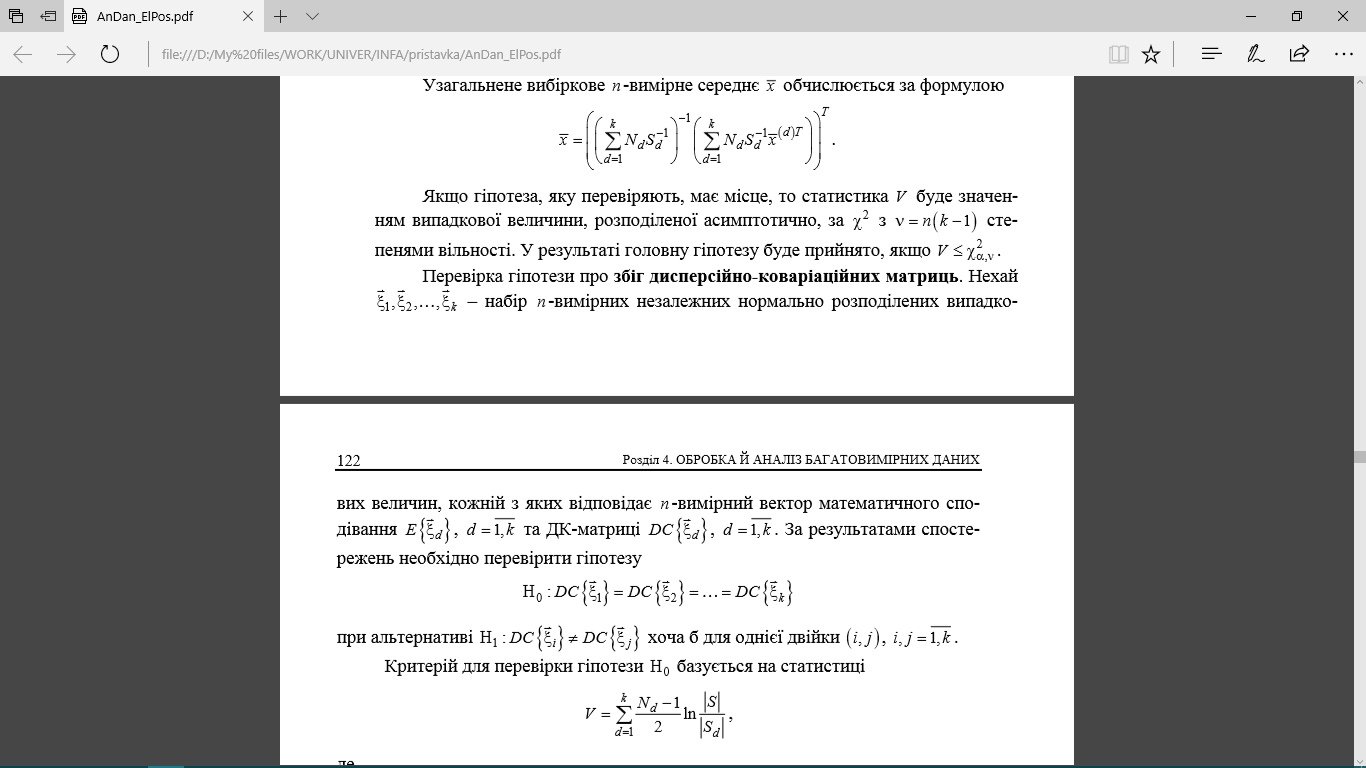
хоча б для однієї двійки . Критерій перевіряється за статистичною характеристикою

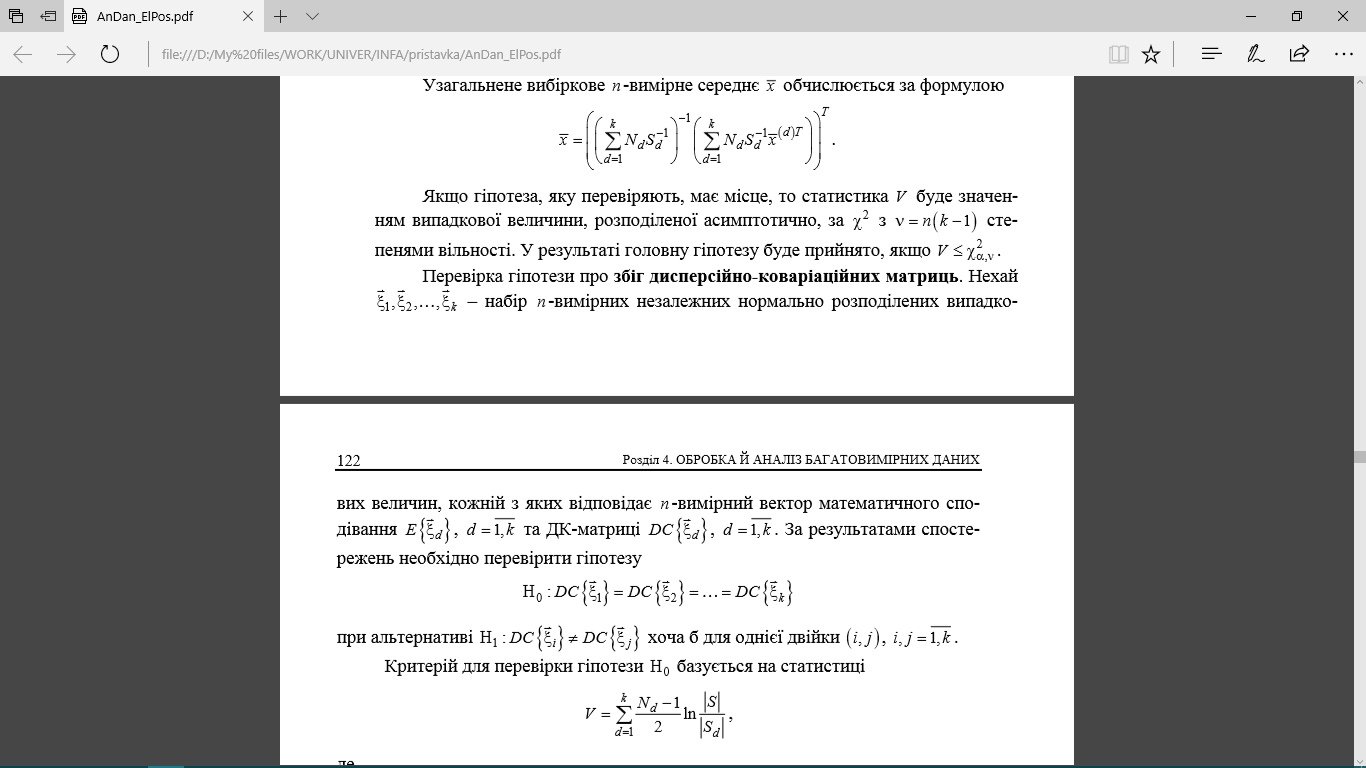
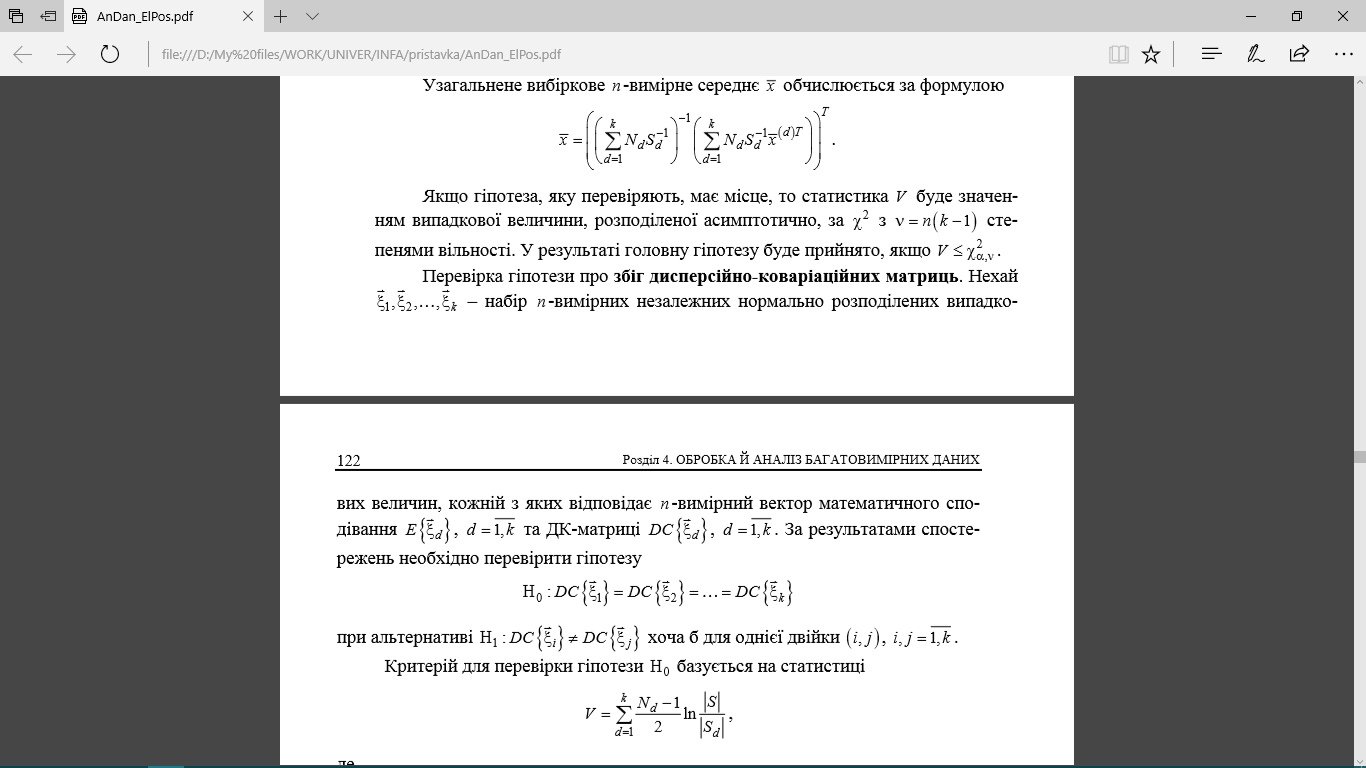


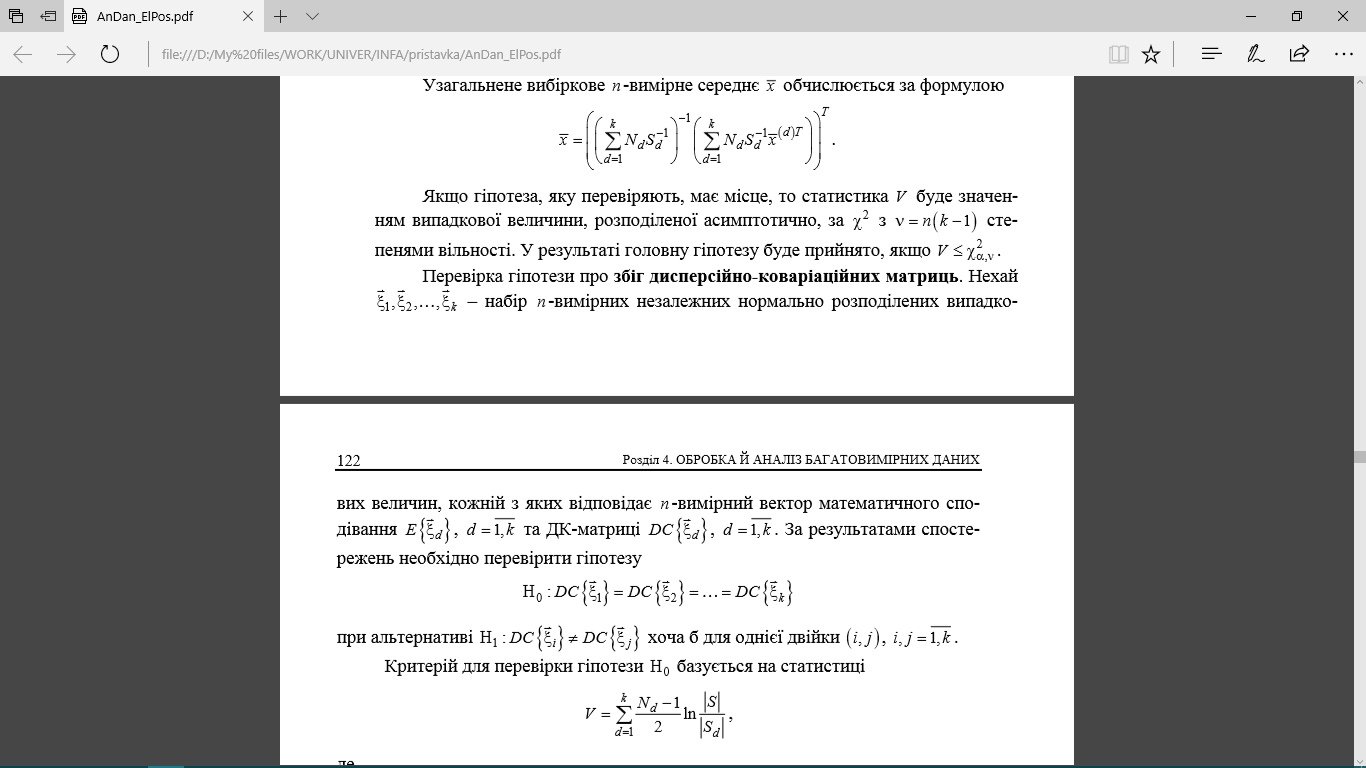
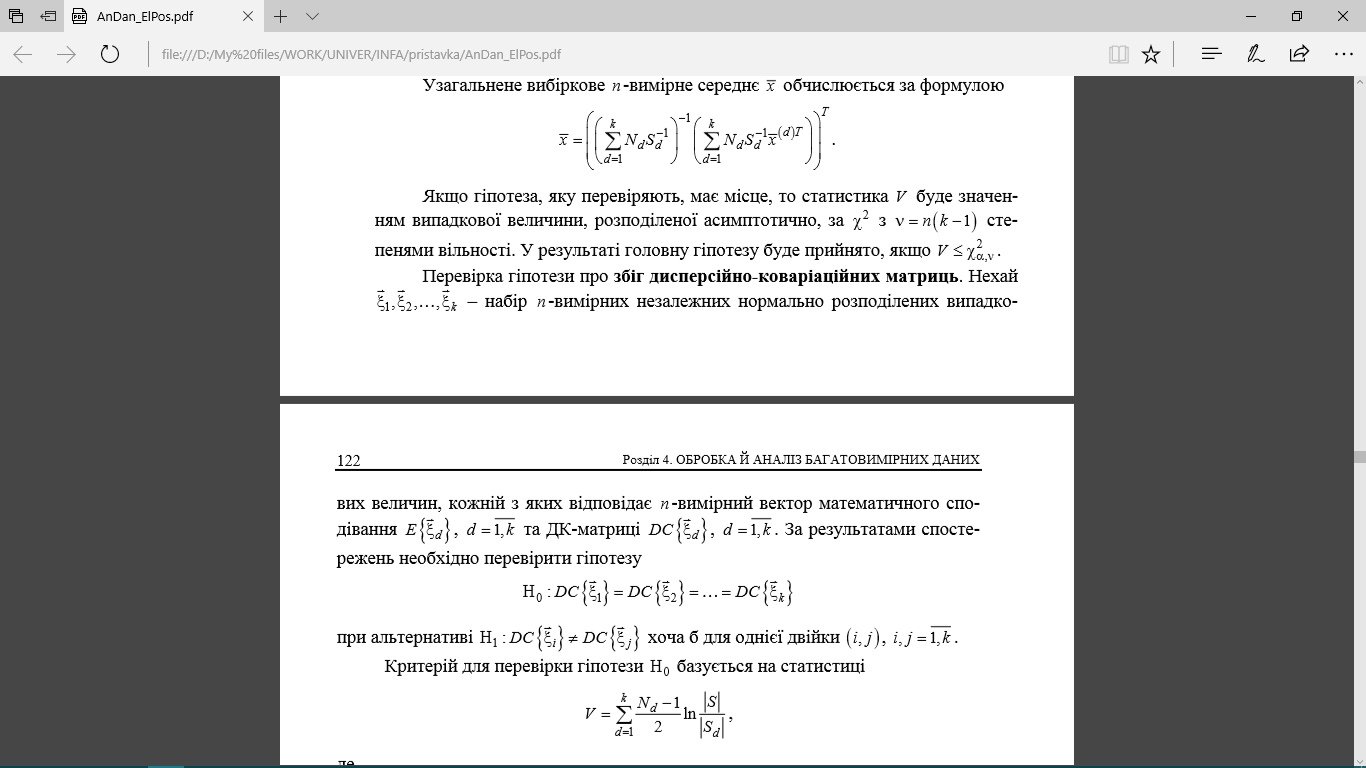
, де

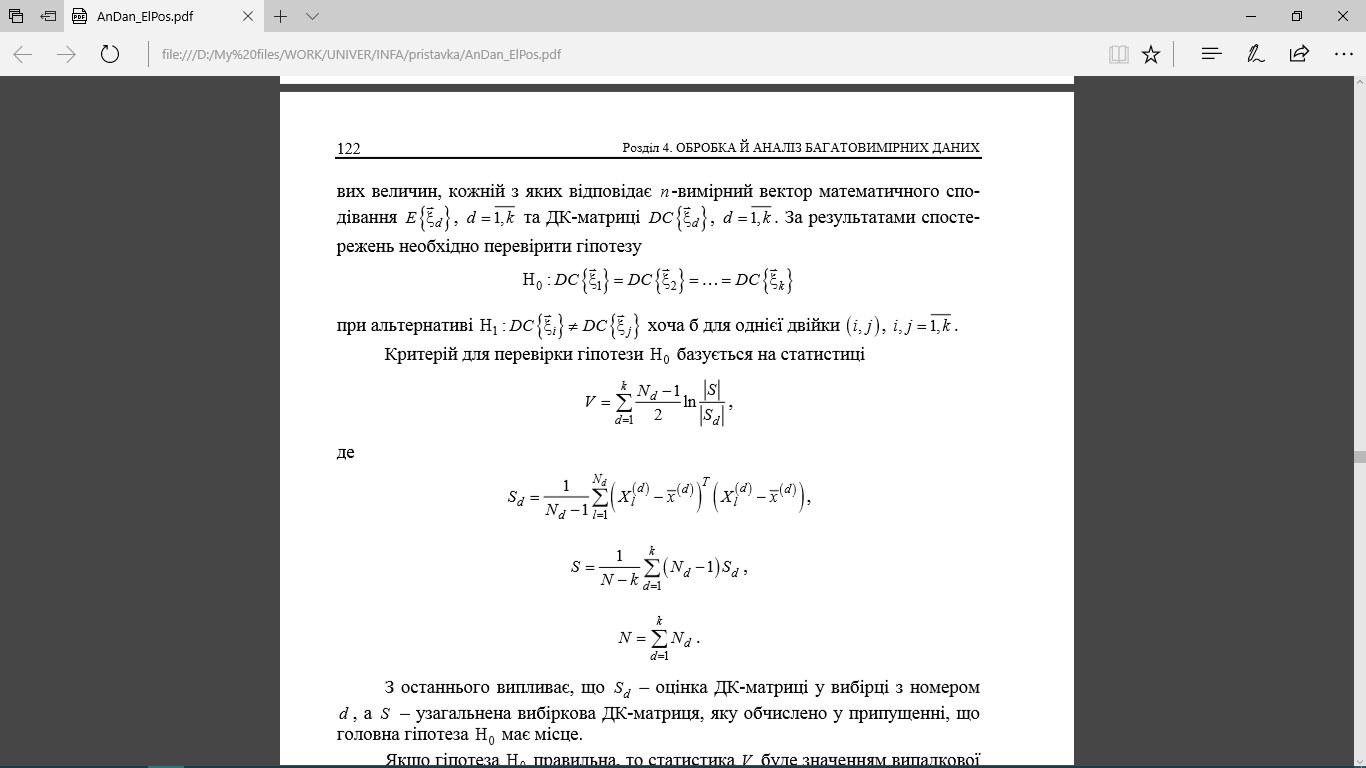


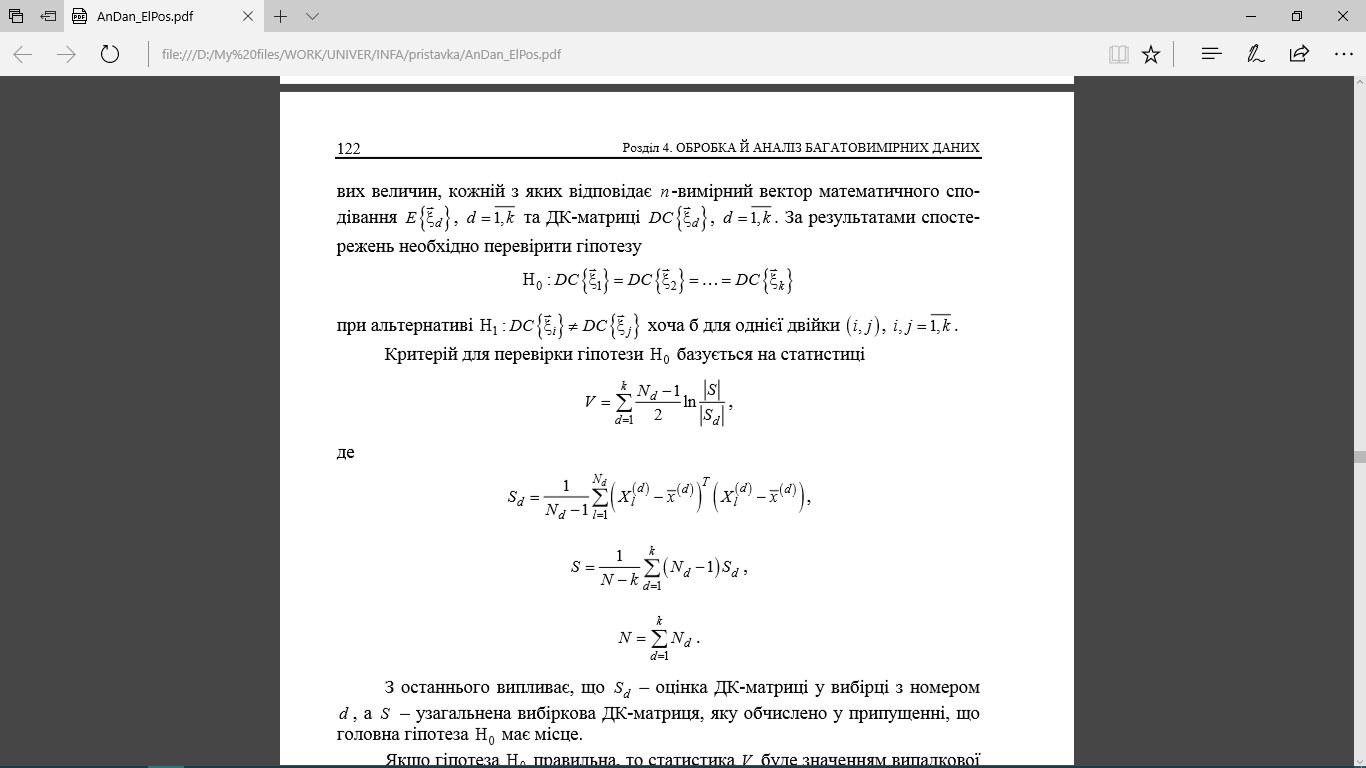
Узагальнене вибіркове n-вимірне середнє x обчислюється за формулою

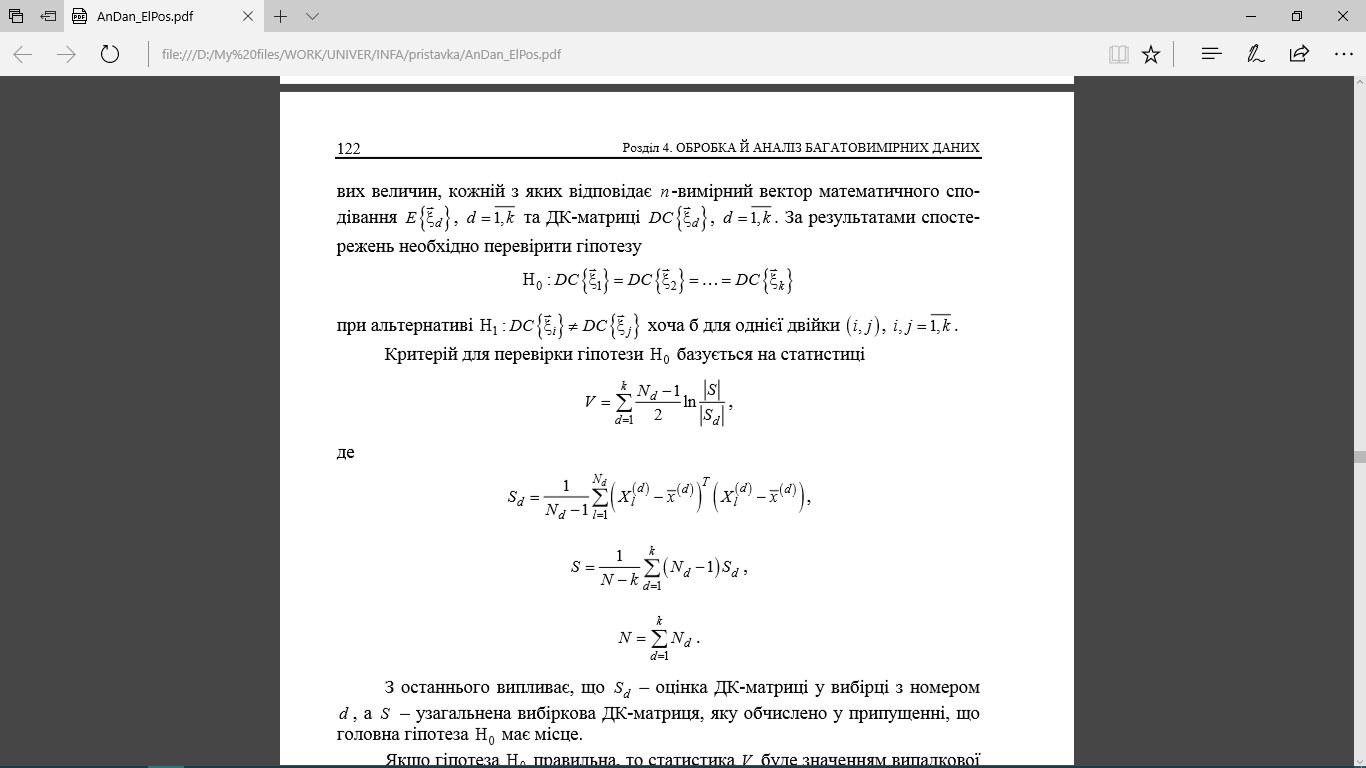


Якщо гіпотеза, яку перевіряють, має місце, то статистика V буде значенням випадкової величини, розподіленої асимптотично, за степенями вільності. У результаті головну гіпотезу буде прийнято, якщо . Перевірка гіпотези про збіг дисперсійно-коваріаційних матриць. Нехай

– набір n-вимірних незалежних нормально розподілених випадкових величин, кожній з яких відповідає n-вимірний вектор математичного сподівання  За результатами спостережень необхідно перевірити гіпотезу



при альтернативі 

хоча б для однієї двійки . Критерій для перевірки гіпотези Η0 базується на статистиці

# ПРАКТИЧНА ЧАСТИНА

Після запуску програми, користувачу необхідно обрати файл, або згенерувати вибірку і додати до древовидної структури у лівому кутку головного екрану.

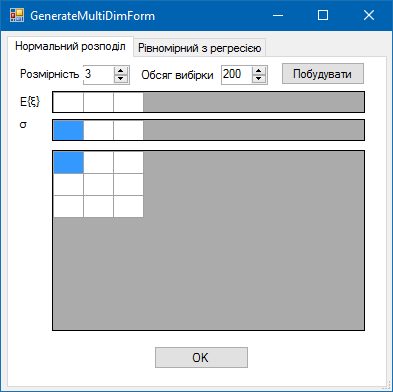


Рис. 1: Форма для генерації вибірки

Всі згенеровані вибірки з'являються у дереві

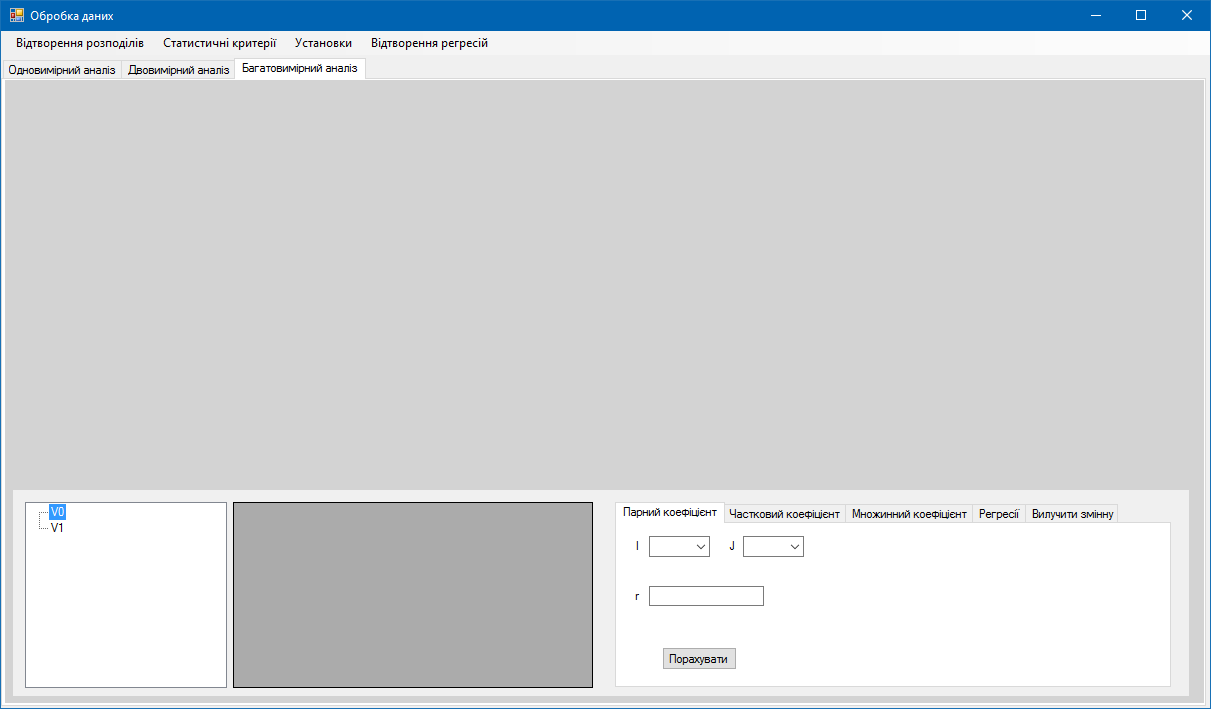


Рис. 2: Додання вибірок до програми,

Потім треба натиснути ПКМ для того, щоб з’явилося контестне меню, та обрати пункт «Аналіз даних на однорідність» Після цього обрати необхідні вибірки

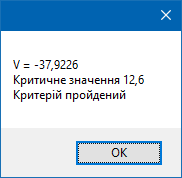
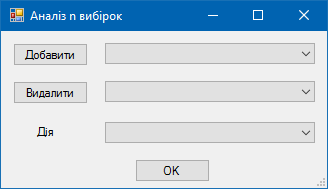


Рис. 2: Обрання вібірок для перевірок гіпотез

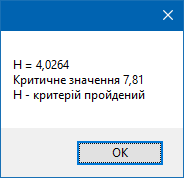
Рис. 3: Відображення результату порівняння 2-х багатовимірних вибірок на збіг середніх

Рис. 4: Відображення результату порівняння 2-х багатовимірних вибірок на збіг середніх

Також можна переглянути матрцю DC для окремої вибірки.

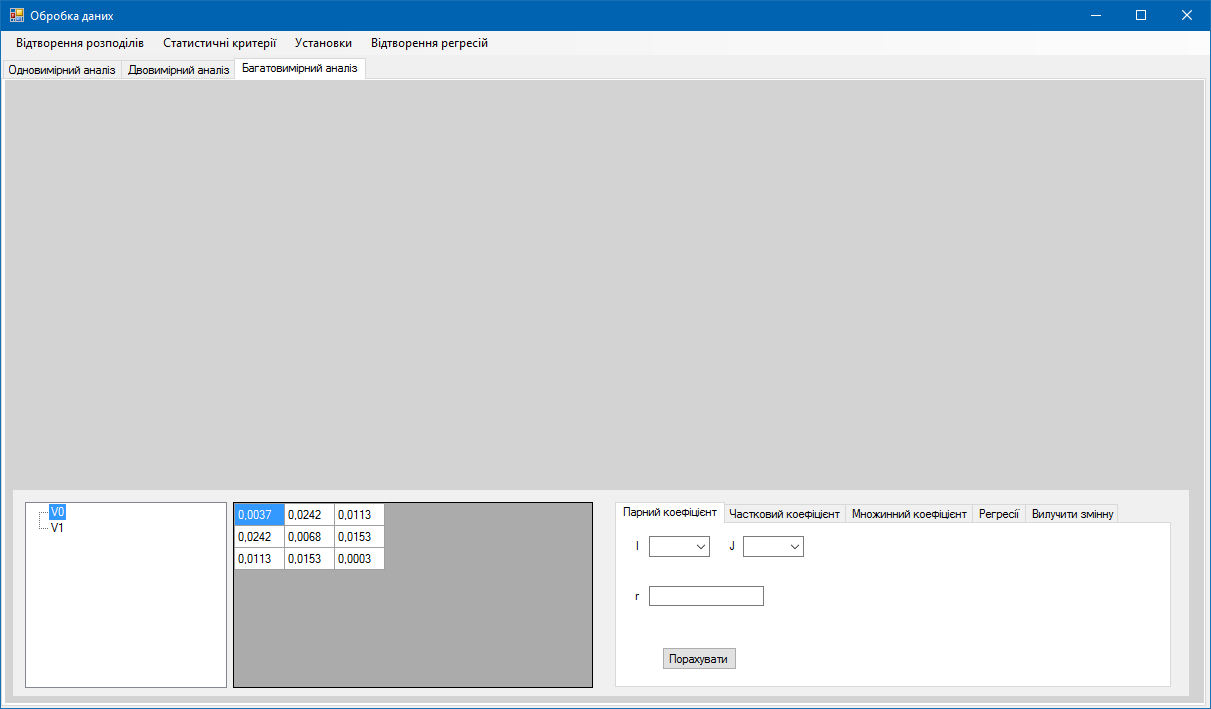
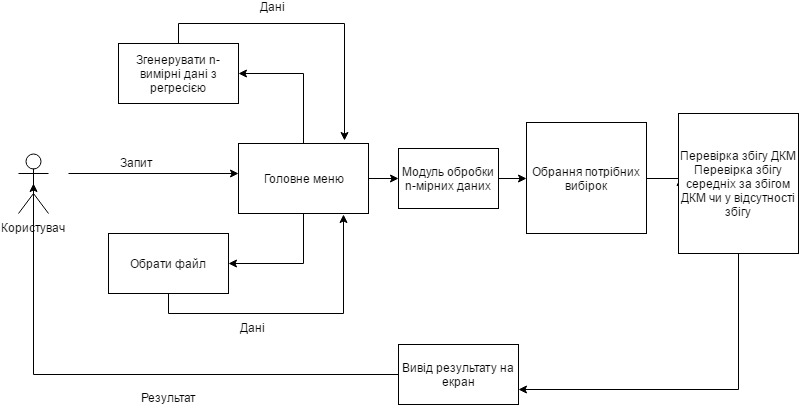


Рис. 5: Матриця DC для обраної вибірки

## UML-діаграма



# ВИСНОВКИ

На даній лабораторній роботі мною було реалізовано первинний статистичний аналіз багатовимірних об’єктів спостережень (знаходження векторів середніх та середньоквадратичних, та побудова дисперсійно-коваріаційної матриці для вибірки;