

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ  
КИЇВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ ТЕХНОЛОГІЙ ТА ДИЗАЙНУ  
ФАКУЛЬТЕТ МЕХІТРОНІКИ ТА КОМП'ЮТЕРНИХ ТЕХНОЛОГІЙ  
КАФЕДРА КОМП'ЮТЕРНИХ НАУК

## **КВАЛІФІКАЦІЙНА РОБОТА**

на тему:

**Комп'ютерна програма для перевірки гіпотези про однорідність багатовимірних показників економічних трендів і процесів.**

Рівень вищої освіти другий (магістерський)  
Спеціальність 122 Комп'ютерні науки  
Освітня програма Комп'ютерні науки

**Виконав:** студент групи МгІТ-2-22  
Халабурський В.В.

**Керівник:** проф. Красницький С.М.

**Рецензент:** доц. Яхно В.М.

КАФЕДРА КОМП'ЮТЕРНИХ НАУК

Спеціальність 122 Комп'ютерні науки

Освітня програма Комп'ютерні науки

**ЗАТВЕРДЖУЮ**

**Завідувач кафедри КНТ  
проф. Володимир Щербань**

“ \_\_\_\_\_ ” \_\_\_\_\_ 2023 року

## **ЗАВДАННЯ**

### **НА КВАЛІФІКАЦІЙНУ РОБОТУ СТУДЕНТУ**

**Халабурському Володимирі Васильовичу**

1. Тема кваліфікаційної роботи: Комп'ютерна програма для перевірки про однорідність багатовимірних показників економічних трендів і процесів.

Науковий керівник роботи: проф. Красницький С.М.

Затверджені наказом вищого навчального закладу від “12” вересня 2023 року № 201-уч

2. Строк подання студентом роботи: «20» листопада 2023 року

3. Вихідні дані до кваліфікаційної роботи: Розробка кафедри комп'ютерних наук. Алгоритми для перевірки гіпотези однорідних багатовимірних показників в економічних та бізнес процесах.

4. Зміст кваліфікаційної роботи (перелік питань, які потрібно опрацювати):

Розділ 1: Базові поняття бізнес-аналітики багатовимірних процесів

Розділ 2: Особливості сучасного програмного забезпечення розв'язання задач багатовимірного статистичного аналізу

Розділ 3: Алгоритм і комп'ютерна реалізація задачі перевірки однорідності багатовимірних статистичних даних

5. Дата видачі завдання\_09.2023р.

## КАЛЕНДАРНИЙ ПЛАН

| № з/п | Назва етапу кваліфікаційної роботи   | Орієнтовний термін виконання | Примітка про виконання |
|-------|--|------------------------------|------------------------|
| 1     | Вступ  | 07.09.2023                   |                        |
| 2     | Розділ 1: Базові поняття бізнес-аналітики багатовимірних процесів  | 20.09.2023                   |                        |
| 3     | Розділ 2: Особливості сучасного програмного забезпечення розв'язання задач багатовимірного статистичного аналізу | 11.10.2023                   |                        |
| 4     | Розділ 3: Алгоритм і комп'ютерна реалізація задачі перевірки однорідності багатовимірних статистичних даних      | 31.10.2023                   |                        |
| 5     | Висновки   | 03.11.2023                   |                        |
| 6     | Оформлення (чистовий варіант)  | 07.11.2023                   |                        |
| 7     | Подача кваліфікаційної роботи науковому керівнику для відгуку (за 14 днів до захисту)                            | 15.11.2023                   |                        |
| 8     | Перевірка кваліфікаційної магістерської роботи на рецензування (за 12 днів до захисту)                           | 17.11.2023                   |                        |
| 9     | Подання кваліфікаційної роботи на наявність ознак плагіату (за 10 днів до захисту)                               | 17.11.2023                   |                        |
| 10    | Подання кваліфікаційної роботи на затвердження завідувачу кафедри (за 7 днів до захисту)                         | 22.11.2023                   |                        |

З завданням ознайомлений:

Студент

\_\_\_\_\_ (підпис)

\_\_\_\_\_ (Власне ім'я та ПРІЗВИЩЕ)

Науковий керівник

\_\_\_\_\_ (підпис)

\_\_\_\_\_ (Власне ім'я та ПРІЗВИЩЕ)

## АНОТАЦІЯ

**Халабурський В.В. Тема: Комп'ютерна програма для перевірки про однорідність багатовимірних показників економічних трендів і процесів.**

Дипломна магістерська робота за спеціальністю 122 – Комп'ютерні науки. Київський національний університет технологій та дизайну, Київ, 2023 рік.

Дипломну магістерську роботу присвячено створенню програмного забезпечення, яке використовує методи перевірки однорідності вибірок у багатовимірній статистиці. Саме програма автономно розраховує потрібні ймовірнісні характеристики розподілів, які виступають у ролі тестових показників, уникнувши при цьому необхідності звертатися до конкретних статистичних програм або таблиць.

Дане програмне забезпечення довело свою корисність у вирішенні різноманітних практичних завдань, таких як порівняння результатів різних технологічних процедур або проведення опитувань серед різних соціальних стандартів та верств населення.

*Ключові слова: багатовимірні дані, бізнес-аналітика, матриці дисперсії, розподіл Фішера.*

## ANNOTATION

### **KHALABURSKYI V.V. COMPUTER PROGRAM FOR VERIFYING THE HYPOTHESIS OF THE HOMOGENEITY OF MULTIDIMENSIONAL INDICATORS OF ECONOMIC TRENDS AND PROCESSES**

This software has been created to employ techniques for verifying the uniformity of multidimensional statistics samples. The program autonomously calculates the relevant probabilistic characteristics of the distributions serving as test indicators, eliminating the need for reference to specific statistical programs or tables. This software proves useful in addressing common practical issues, such as comparing outcomes from various technological procedures or conducting surveys across different population segments and social standards.

*Keywords: multidimensional data, dispersion matrices, Fisher distribution.*

## ЗМІСТ

|   |    |
|---|----|
| Вступ.....  | 9  |
| РОЗДІЛ 1 БАЗОВІ ПОНЯТТЯ БІЗНЕС-АНАЛІТИКИ БАГАТОВИМІРНИХ<br>ПРОЦЕСІВ.....  | 11 |
| 1.1. Сутність багатовимірного статистичного аналізу. Можливості застосування<br>багатовимірного статистичного аналізу в бізнес-аналітиці.....             | 11 |
| 1.2. Історичні аспекти використання багатовимірного статистичного аналізу.<br>Методи багатовимірного статистичного аналізу.....                           | 16 |
| 1.3. Особливості обробки багатовимірних статистичних даних. Види простору<br>ознак. Етапи дослідження за допомогою багатовимірного статистичного аналізу. | 19 |
| 1.4. Принципи перевірки статистичних гіпотез. Алгоритми проведення<br>статистичного тестування.....   | 31 |
| Висновки до 1 розділу.....  | 42 |
| РОЗДІЛ 2 ОСОБЛИВОСТІ СУЧАСНОГО ПРОГРАМНОГО ЗАБЕЗПЕЧЕННЯ<br>РОЗВ'ЯЗАННЯ ЗАДАЧ БАГАТОВИМІРНОГО СТАТИСТИЧНОГО АНАЛІЗУ...                                     | 43 |
| 2.1. Аналіз сучасного стану на ринку програмного забезпечення статистичної<br>обробки даних.....  | 43 |
| 2.2. Дослідження характеристик програмних засобів для математичної і<br>статистичної обробки даних.....   | 45 |
| 2.3. Висновки до 2 розділу.....   | 75 |
| РОЗДІЛ 3 АЛГОРИТМ І КОМП'ЮТЕРНА РЕАЛІЗАЦІЯ ЗАДАЧІ ПЕРЕВІРКИ<br>ОДНОРІДНОСТІ БАГАТОВИМІРНИХ СТАТИСТИЧНИХ ДАНИХ.....  | 76 |
| 3.1. Розрахункові співвідношення та програмне забезпечення.....   | 76 |
| Висновки до 3 розділу.....  | 97 |
| Висновки.....   | 98 |
| Список використаних джерел.....   | 99 |

## ВСТУП

В сучасних умовах, коли рішення приймаються базуючись на підставі стохастичної, неповної та частково відомої інформації, використання методів багатовимірного статистичного аналізу є вкрай необхідним. Соціально-економічні та технологічні процеси і явища залежать від значної кількості параметрів, які їх характеризують. Це обумовлює деякі труднощі, пов'язані з виявленням структури взаємозв'язків даних параметрів. Сучасні методи багатовимірного аналізу даних використовуються для розв'язання задач, пов'язаних з дослідженням поведінки індивідуума, родини, іншої соціально-економічної одиниці як представника достатньо великої сукупності об'єктів і суб'єктів. Спектр методів багатовимірного статистичного аналізу сьогодні досить широкий. Наприклад, вивчення взаємозв'язків у багатовимірних сукупностях можна виконувати використовуючи кореляційний та регресійний аналіз. Для оцінки тісноти зв'язку між системами показників використовують канонічні кореляції.

Методи багатомірної класифікації призначаються для розподілу сукупності об'єктів на визначені однорідні клатстерні групи. Кожний з об'єктів отримує характеристику великою кількістю різноманітних стохастично зв'язаних ознак. Для того щоб розв'язати задачу класифікації застосовують кластерний і дискримінантний аналіз.

Присутність множини вихідних ознак, що характеризують багатовимірні об'єкти, викликає необхідність визначати найбільш істотні та вивчати значно менший набір показників. Для відбору, групи змінних і ранжування використовують методи таксономії та вибору репрезентантів. Розв'язання задач зниження розмірності простору ознак забезпечують методи факторного, компонентного аналізу, багатовимірного шкалювання. Ці методи розкривають

існуючі закономірності, які безпосередньо не виявляються, за допомогою факторів, головних компонент і шкал. Стиснення інформації здійснюється за рахунок кількості факторів або головних компонент, яка значно менша, ніж кількість існуючих вихідних ознак.

Теперішні ринково-економічні умови потребують від суб'єктів господарювання вміння аналізувати багатовимірні процеси та предметно використовувати методи і моделі багатовимірного аналізу даних для прийняття ефективних і надійних управлінських рішень.

Розробка програмного забезпечення, що реалізує методи перевірки однорізності вибірок багатовимірних статистичних даних являється важливою та пріоритетною для сучасної аналітики бізнес процесів. В наш час, обчислення належних імовірнісних характеристик розподілів, що виконують роль індикаторів перевірки (розподілів Фішера) виконуються за допомогою програм, в результаті, необхідності в переході до спеціальних статистичних програм або таблиць більше немає. Подібні програмні рішення можуть застосовуватися для розв'язання доволі часто виникаючих на практиці задач пов'язані з порівнянням різноманітних технологічних процесів, опитувань та соціальних стандартів.



## РОЗДІЛ 1. БАЗОВІ ПОНЯТТЯ БІЗНЕС-АНАЛІТИКИ БАГАТОВИМІРНИХ ПРОЦЕСІВ

### 1.1. Сутність багатовимірного статистичного аналізу. Практики та методи застосування багатовимірного статистичного аналізу в бізнес-аналітичних процесах.

Багатовимірний статистичний аналіз (БСА) являється інтелектуальним інструментом дослідження. Постійно зростаючий інтерес до БСА пояснюється першочергово його можливостями у моделюванні та відображенні реальних процесів і явищ, що найчастіше мають природу що складається з багатьох ознак.

Статистика, математика та інші точні науки не можуть розвиватися без базових знань з обробки багатовимірних даних. Усі сучасні передові розробки, які описують проблеми нечітких множин, моделювання катастроф, розпізнаванню образів за допомогою штучного інтелекту (ШІ), сценарного прогнозування тощо, допускають багатовимірне подання об'єктів за якими ведеться спостереження. Основні можливості БСА показано на рис. 1.1.



Рис. 1.1. Основні можливості БСА

БСА – це сукупність формалізованих статистичних методів, які базуються на поданні результативної інформації в багатовимірному геометричному просторі та дозволяють визначати приховані (латентні), але об'єктивно існуючі закономірності в організаційній структурі та тенденціях розвитку соціально-економічних явищ і процесів.

Застосування багатовимірного статистичного аналізу в економічних процесах  
Торговельні та промислові підприємства:

- оцінювання класу інвестиційної привабливості;
- банкрутства та прогнозування його класів;
- економічна безпека та її аналіз на підприємстві;
- оцінювання та позиціонування рівня економічного розвитку підприємства.

Роздрібна торгівля:

- аналіз купівельної спроможності та корзини для виявлення товарів, які покупці бажають купувати в першу та другу чергу, розроблення напрямків стратегії створення запасів товарів на складах і найефективніших способів їх розміщення в торгових залах супермаркетів та торгових центрів;
- дослідження характеру потреб та поведінокві характеристики різноманітних категорій споживачів з певною поведінковою моделлю (наприклад, ті, які купують товари відомих брендів або дизайнерів, а також відвідують щосезонні розпродажі), для розроблення стратегії та окремих заходів із просування товарів серед насиченого ринку споживачів.

Банківська індустрія:

- визначення та виявлення шахрайства по користуванню кредитними картками за допомогою всебічного аналізу минулих транзакцій;

- сегментація клієнтів підрозділами відділу маркетингу, фактичне прогнозування змін в рамках категорій споживачів, моделі прогнозів та цінності своїх клієнтів.

Страховання:

- сканування та оцінка стереотипів у заявах про виплату страхового відшкодування;
- аналіз ризиків шляхом виявлення та поєднань факторів, пов'язаних зі зверненнями по виплатам.

Онлайн інтернет-продажі:

- розроблення ефективних методів та підходів для обробки, очищення та агрегації дослідження даних.

## **1.2.Історичні аспекти використання багатовимірного статистичного аналізу. Методи багатовимірного статистичного аналізу.**

Багатопараметричний опис явищ, об'єктів і процесів достатньо широко виствітлюється в роботах таких вчених, як:

|  |  |
|--|--|
| Ч. Дарвін (60-ті рр. XIX ст., англійський натураліст)  | селекція видів та визначення чинників еволюції органічного світу   |
| Д.І. Менделєєв (60–70 рр. XIX ст.)                     | систематизація якісних характеристик хімічних елементів і процесів |
| А.П. Шлікевич, І.Ф. Анненський та ін. (початок XX ст.) | класифікація земельних господарств                                 |

Історію БСА як наукового самостійного напрямку розвитку статистичної теорії пов'язують з публікаціями на початку XX ст. англійських вчених У.

Спірмена та К. Пірсона, присвячених основам побудови алгоритмів стиснення статистичних даних. Фундаментальні ознаки теорії БСА заклали такі вчені, як Л.Л. Терстоун, Р. Хорст, Г. Кайзер, Л.Р. Такер, Т. Келлі, Г. Харман, М. Жамбю, Р.В. Хемінг, Л. Заде, Р. Тріонон, Р. Сокал, Р. Фішер та ін. (рис. 1.2).



Рис 1.2. Вчені-теоретики, які заклали фундаментальні принципи теорії БСА

Методи багатовимірного статистичного аналізу будуються на статистичних методах, таких як теорія ймовірностей, математична статистика, загальна теорія статистики та на методах вищої математики, такі як аналітична геометрія, матрична алгебра, багатомірний математичний аналіз. В свою чергу вони використовують методи ймовірнісного аналізу даних і методи логіко-геометричного напрямку (рис. 1.3).



Рис 1.3. Методи багатовимірного статистичного аналізу

Отже, можемо прийти до висновку, що методи багатовимірного статистичного аналізу дозволяють розв'язувати різноманітні задачі дослідження багатовимірних процесів:

- групування
- стиснення даних
- моделювання складних систем
- оцінювання параметрів багатовимірних сукупностей та ін.

Такий підхід дозволяє використовувати отримані результати для прийняття ефективних управлінських рішень.

### 1.3. Особливості обробки багатовимірних статистичних даних. Види простору ознак. Етапи дослідження за допомогою багатовимірного статистичного аналізу.

Обробка багатовимірних статистичних даних має власні специфічні особливості: застосування практичних методів багатовимірного статистичного аналізу вимагає творчого і креативного підходу для розв'язання важких аналітичних задач, по причині обробки багатовимірних сукупних даних. Для методів багатовимірного статистичного аналізу характерна глибока формалізація та складна математична та логічна конструкція, їх практичне застосування супроводжується використанням обчислювальної техніки (рис. 1.4).



Рис 1.4. Особливості застосування методів багатовимірного статистичного аналізу

Найбільш структурованою формою подання вихідних даних про багатовимірний об'єкт є матрична.

Види матриць:

- Матриця значень аналітичних ознак ( $x_j$ )

| Об'єкт | $X_1$    | $X_2$    | $X_3$    | $X_4$    | ... | $X_m$    |
|--------|----------|----------|----------|----------|-----|----------|
| $n_1$  | $x_{11}$ | $x_{12}$ | $x_{13}$ | $x_{14}$ | ... | $x_{1m}$ |
| $n_2$  | $x_{21}$ | $x_{22}$ | $x_{23}$ | $x_{24}$ | ... | $x_{2m}$ |
| $n_3$  | $x_{31}$ | $x_{32}$ | $x_{33}$ | $x_{34}$ | ... | $x_{3m}$ |
| ...    | ...      | ...      | ...      | ...      | ... | ...      |
| $n_n$  | $x_{n1}$ | $x_{n2}$ | $x_{n3}$ | $x_{n4}$ | ... | $x_{nm}$ |
|        |          |          |          |          | ... |          |

- Матриця теоретичних відстаней між об'єктами ( $n_j$ )

Види простору ознак:

- з нульовою розмірністю (об'єкти не мають характеристик);
- одномірний простір ознак (об'єкти відображаються значеннями однієї ознаки);
- багатовимірний простір (об'єкти подані значеннями двох і більше ознак).

Приклад одновимірного простору ознак (рис. 1.5).

| Підприємство<br>№п/п | Середньостатистичний<br>рівень виробітку одного<br>робітника, дол. США<br>(X) |
|----------------------|---|
| <b>1</b>             | <b>6</b>  |

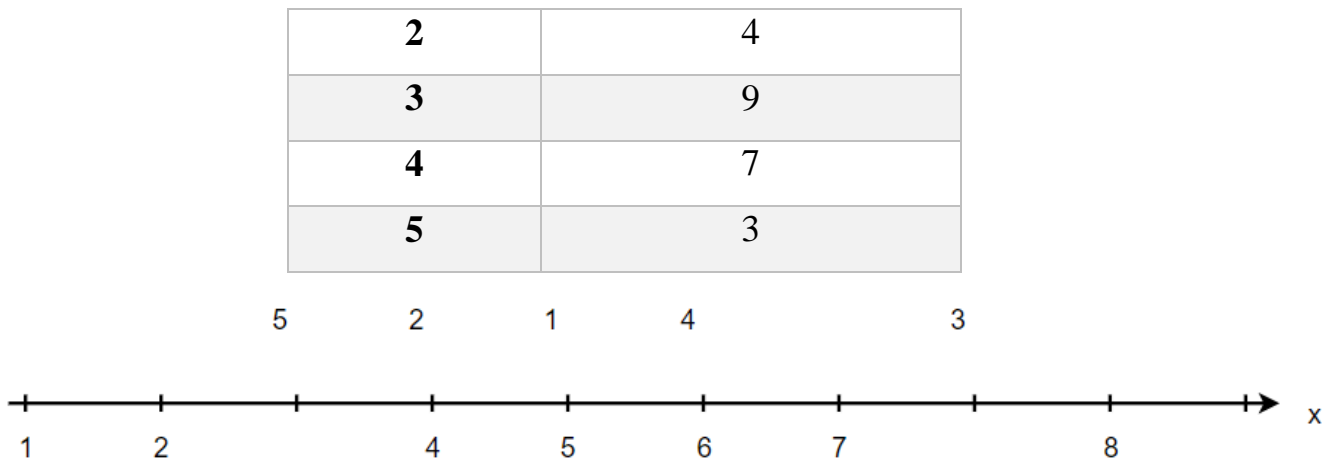


Рис 1.5. Одновимірний простір ознак

Приклад двовимірного простору ознак (рис. 1.6).

| Підприємство<br>№п/п | Середньостатистичний<br>рівень виробітку одного<br>робітника, дол. США<br>(X) | Середній стаж роботи<br>робітника на<br>підприємстві, років (X2) |
|----------------------|---|--|
| <b>1</b>             | 6   | 5  |
| <b>2</b>             | 4   | 7  |
| <b>3</b>             | 9   | 12   |
| <b>4</b>             | 7   | 14   |
| <b>5</b>             | 3   | 11   |



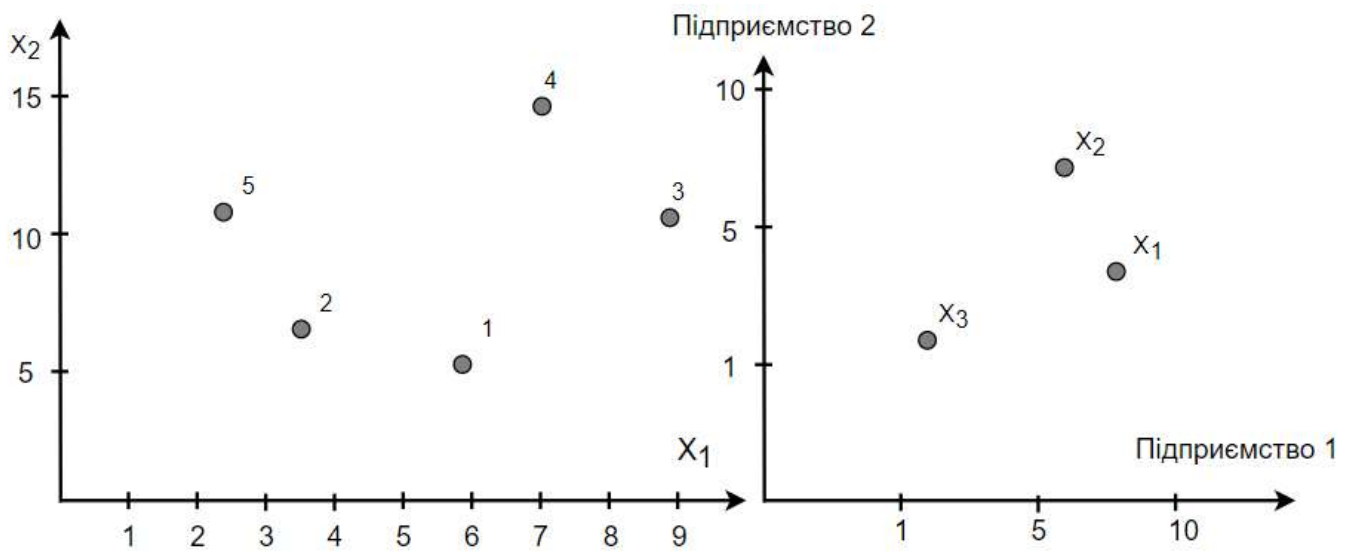


Рис 1.6. Двовимірний простір ознак

Приклад тривимірного простору ознак (рис. 1.7).

| Підприємство<br>№п/п | Середньостатистичний<br>рівень виробітку<br>одного робітника, дол.<br>США (X) | Середній стаж<br>роботи робітника на<br>підприємстві, років<br>(X2) | Середній рівень<br>класифікації<br>робітників за<br>тарифним<br>розрядом (X3) |
|----------------------|---|---|---|
| <b>1</b>             | 6   | 5   | 1,2   |
| <b>2</b>             | 4   | 7   | 1,9   |
| <b>3</b>             | 9   | 12  | 3,5   |
| <b>4</b>             | 7   | 14  | 2,7   |
| <b>5</b>             | 3   | 11  | 2,6   |

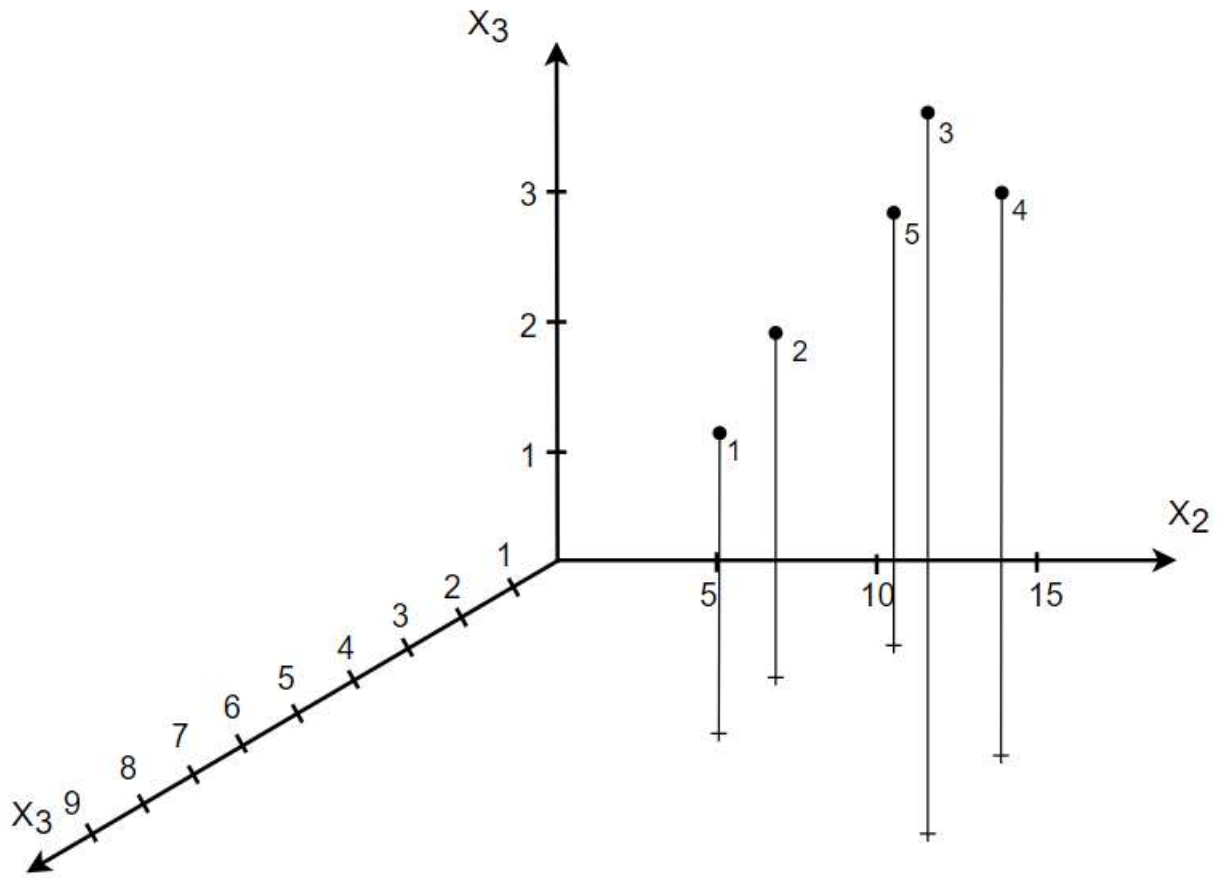


Рис 1.7. Тривимірний простір ознак

Сучасна економічна ситуація вимагає, щоб суб'єкти господарювання були здатні аналізувати складні процеси та використовувати результати аналізу даних для прийняття продуктивних управлінських рішень. При виборі конкретної процедури прийняття рішень або механізму управління, орган управління повинен передбачати, як агенти - суб'єкти управління, будуть реагувати на свої дії, враховуючи цілі організації. (рис. 1.8).



Рис. 1.8. Загальна схема процесу формувань рішень

Експериментування в реальному житті, шляхом випробовування різних управлінських дій і вивчення реакції підлеглих, часто є неефективним та практично неможливим. У цьому випадку моделювання стає важливим інструментом. Моделювання - це метод дослідження, який включає в себе створення та аналіз моделей, які є аналогами об'єктів дослідження. За допомогою адекватної моделі можна аналізувати реакцію системи управління (це відбувається на етапі аналізу, рис. 1.8), а потім на основі цього аналізу вибирати і впроваджувати на практиці ті управлінські впливи, які призводять до бажаної реакції. Наявність моделей та механізмів управління корисна як для органів управління (оскільки це дозволяє передбачити поведінку суб'єктів управління), так і для самих суб'єктів (оскільки це робить поведінку управлінського органу передбачуваною). Зменшення невизначеності завдяки використанню механізмів управління є важливою властивістю будь-якої організації як соціального інституту.

Вивчення економічних об'єктів будь-якого виду включає розкриття якісних і кількісних закономірностей через створення моделей для аналізу багатовимірних об'єктів. Модель може бути сприймана як образ реального об'єкта (або процесу), представлений у матеріальній або ідеальній формі (тобто відтворений за допомогою знакових мов), що відтворює важливі властивості модельованого об'єкта (або процесу) та допомагає в його аналізі під час дослідження.

Проблеми можна класифікувати наступним чином:

- Кількісні проблеми, які вимагають роботи з числами та обчислень.
- Якісні проблеми, пов'язані з аналізом майбутніх або невизначених ресурсів і їх властивостей.
- Слабоструктуровані проблеми, які характеризуються тим, що не всі елементи, їх властивості та зв'язки відомі повністю.

Ми далі розглянемо економіко-математичне моделювання за допомогою методів багатовимірного аналізу, тобто створення математичних моделей соціально-економічних систем. Існують різні види моделювання:

- Вербальне моделювання використовує мовні засоби для створення моделей.
- Геометричне моделювання передбачає створення моделей на основі макетів або об'єктних моделей, які передають просторові форми об'єкта.
- Фізичне моделювання використовується для вивчення фізико-хімічних, технологічних та інших процесів, що відбуваються в оригіналі.
- Інформаційне моделювання включає в себе використання різних засобів, таких як схеми, графіки, креслення, формули, рівняння та нерівності.

Логіко-математичне моделювання є важливою складовою інформаційного моделювання і використовує математичний апарат для аналізу системи. При побудові економіко-математичних моделей застосовується метод моделювання за

принципом "чорного ящика", де дослідник не знає внутрішнього механізму системи, але може аналізувати її за вхідними та вихідними характеристиками (рис. 1.9).

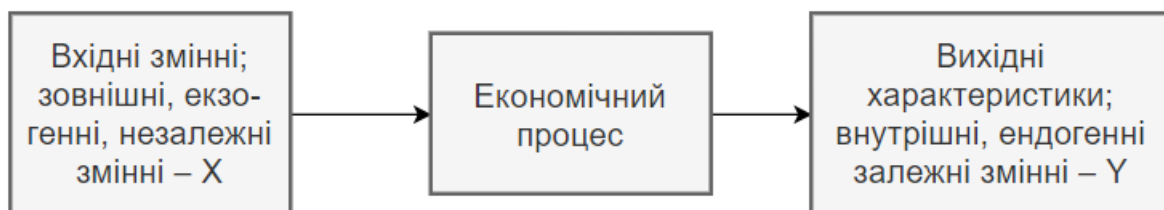


Рис.1.9. Дослідження системи за принципом «чорного ящика»

Для вивчення соціально-економічних систем і розкриття їх причинно-наслідкових зв'язків, дослідник повинен виділити вхідні та вихідні характеристики системи. Це передбачає такі етапи дослідження за допомогою методів багатовимірного статистичного аналізу:

**Постановка задачі:** Включає в себе опис предметної області об'єктів та визначення обсягів виділених ресурсів, таких як час і трудові витрати.

**Вибір методів:** Визначення набору методів багатовимірного статистичного аналізу та порядку їх використання.

**Збирання даних:** Включає в себе визначення способів збирання інформації та форм її подання.

**Аналіз даних:** Перевірка однорідності вибірки, відповідність законам розподілу, виявлення грубих помилок тощо.

**Уточнення математичної постановки задачі:** Визначення можливості застосування раніше відібраних методів.

Реалізація: Проведення обчислень та реалізація за допомогою програмного забезпечення математичного інструментарію.

Оцінювання адекватності моделі: Визначення несуперечності математичних результатів і економічних висновків.

Важливим принципом управління економічними об'єктами є принцип системності. Системи можуть бути подані нескінченним числом структурних та функціональних інваріантів, які відображають взаємозв'язки між різними процесами, які відбуваються у системі, такі як економічні, соціальні, екологічні, демографічні тощо.

Ця система описується якісними і кількісними параметрами, які є основою для математичного опису процесів функціонування системи. Параметри визначають мову опису системи, і шляхом їх формалізації вони стають незалежними змінними математичного опису процесу функціонування системи.

Точність моделі і ймовірність досягнення позитивних результатів від її використання великою мірою залежать від того, наскільки повні та обгрунтовані принципи моделювання та системності. Основні принципи моделювання включають в себе принцип інтегратизма, який підкреслює важливість взаємодії між частинами системи та виникнення нових властивостей у системі, а також детермінованість просторових і функціональних зв'язків.

Чим краще деталізовано і обгрунтовано принципи моделювання, тим краща і вища точність моделі, а досягнення позитивних результатів від її застосування підвищує імовірність в багато разів. Нижче наведені основні принципи, які відіграють найбільшу роль у моделюванні:

**Принцип інтегратизма** полягає в тому, що окремі частини одного цілого характеризуються сукупністю трьох елементів:

1. виникненням взаємодіючих систем — зв'язки між частинами цілого;
2. втрата деяких властивостей частини зі входженням у ціле;
3. поява нових властивостей у цілого, обумовлених властивостями складових.

Принцип інтегратизма визначає, що коли розглядаємо окремі частини складної системи, вони взаємодіють між собою, при цьому деякі з їхніх властивостей можуть змінюватися через цю взаємодію, і в результаті виникають нові властивості, які характерні лише для цієї цілісної системи. Важливою умовою є логічна організація і зв'язок між цими частинами, де кожна частина стає неодмінною частиною цілої системи, яка внутрішньо є цілісною і взаємодіє внутрішньо. Цей підхід визначається принципом інтегратизма, хоча і суперечить поглядам У. Ешбі, який розглядає загальну теорію систем як загальну теорію спрощення.

**Принцип невизначеності** вказує на те, що на екстремальних точках економічні процеси є рухливими та неоднозначними. Вони постійно змінюються у часі, і навіть якщо нам вдається зафіксувати певну якість або властивість процесу, вона залишається актуальною лише для конкретного моменту часу та обставин. Інакше кажучи, на мікрорівні економічні процеси повинні бути розглянуті з урахуванням випадкової зміни факторів.

При цьому принцип невизначеності показує, що існують рівні факторів, на яких невеликі зміни не впливають на стан системи. Однак, з ростом складності моделі системи та глибшим аналізом, рішення завдань стає більш невизначеним і віддаляється від практичної застосовності.

**Принцип інваріантності** полягає в тому, що модель системи являється інваріантною для будь-яких організаційних форм виробництва, а зміна умов не має впливати на сутність моделі.

Принцип головних видів діяльності полягає в тому, що різні системи мають «схожі» види діяльності (розподів, регулювання, управління і т. д.), які визначаються як стандартні. Вони можуть бути незмінні на певному проміжку часу, в той самий час їх можна описати деякими схожими за принципами моделями.

Для ефективного моделювання об'єктів та подання їх у вигляді систем потрібно брати до уваги властивості економічних систем — а саме:

- Емерджентність - це властивість системи, яка проявляється найяскравіше в ідеї цілісності. Це означає, що в економічній системі існують характеристики або явища, які не можна пояснити або присвоїти жодному окремому елементу системи, взятому окремо від системи. Емерджентність виникає завдяки взаємодії між різними елементами системи, яку називають синергією. Ця синергія дозволяє досягнути загального ефекту, який перевищує суму ефектів окремих елементів системи, коли вони працюють незалежно один від одного.
- Подільність — основний цілісний об'єкт має бути зображений поділенням на малі елементи;
- Ідентифікованість — кожен окремий елемент системи може бути відділений від інших складових як індивідуальна частина;
- Стабілізація — система здійснює відновлення своїх елементів за рахунок їхнього регулювання;
- Спостережність — як правило усі без винятку входи та виходи системи, також можуть контролюватися дослідником, або в найкращому випадку дослідник може за ними просто спостерігати та аналізувати;
- Неможливість ізолювати процеси, які відбуваються в різноманітних економічних системах, також явища навколишнього середовища, для того щоб спостерігати та досліджувати їх у природньому вигляді;



- Стійкість — дана система повинна нормально функціонувати та бути стійкою до ймовірних сторонніх факторів, які можуть впливати на неї та дестабілізувати її;
- Розмаїтість — для кожного елемента системи властиве поведження та стан, відмінні від поведження та стану інших елементів і системи в цілому;
- Невизначеність — дослідник в один і той самий час не може фіксувати всі властивості та відношення елементів системи. Для того щоб виявити їх і проводиться системне дослідження;
- Нетотожність відображення — основна знакова система дослідника, в свою чергу вона відмінна від знакової системи прояву властивостей об'єктів і їхніх відношень. Втрата аналітичних даних визначає нетотожність системи досліджуваному об'єкту;
- Динамічність економічних процесів полягає в зміні структури та параметрів економічних систем під впливом зовнішніх факторів (середовища);
- Випадковість і невизначеність у розвитку економічних явищ — подібні закономірності економічних процесів виявляються на базуючись на випадкових величинах та факторах, їхні параметри розподілу встановлюються з деякою приблизною ймовірністю;
- Активна реакція на нові фактори, які з'являються — це насамперед здатність соціально-економічних систем до активних, та не завжди передбачуваних дій, яке залежить від відношення системи до цих факторів.
- Адаптація — стан коли система зберігає рухливу рівновагу та стійкість до сторонніх дестабілізаційних факторів. Вона постійно піддається таким впливам, шляхом перебудови внутрішньої структури, а також функцій окремих елементів;
- Масовий характер економічних явищ і процесів — закономірність економічних процесів, які не виявляються на підставі невеликої кількості спостережних даних;

## **1.4. Принцип перевірки статистичних гіпотез. Алгоритми проведення статистичного тестування**

### **Принципи перевірки статистичних гіпотез**

Статистичні гіпотези в галузі статистики грають важливу роль, оскільки вони дозволяють перевіряти припущення стосовно генеральної сукупності на основі вибірки. Важливо зрозуміти, що метою перевірки статистичної гіпотези є не її доведення, а лише визначення, чи можна її підтвердити чи спростувати. Процес перевірки гіпотези пов'язаний з можливістю зробити помилку, яку зазвичай поділяють на помилки першого та другого роду.

Статистична гіпотеза (гіпотеза) - це заявлення або припущення, що стосується генеральної сукупності і перевіряється на основі вибірки. Гіпотеза може бути параметричною, коли перевіряється припущення щодо параметрів розподілу, або непараметричною, коли перевіряється гіпотеза про форму розподілу, яка не визначена конкретними параметрами.

Перевірку статистичної гіпотези можна проводити за допомогою параметричних та непараметричних методів. Зазвичай на практиці існують два типи завдань щодо перевірки гіпотез: перше завдання пов'язане з перевіркою наявності статистично важливих відмінностей між параметрами статистичних сукупностей, такими як середні, дисперсії, коефіцієнти кореляції тощо. Другий тип завдань пов'язаний з перевіркою гіпотез щодо істотності відмінностей у розподілі вибірових даних від теоретичного розподілу, який часто є нормальним, а також з перевіркою однорідності складу кількох сукупностей та інших схожих завдань.

Давайте розглянемо деякі достатньо важливі передумови й особливості перевірки статистичних гіпотез задач першого типу, котрі пов'язані із застосуванням параметричних критеріїв і припущення нормального розподілу в

генеральній сукупності. Набір конкретних заходів з перевірки гіпотези залежить від наступних умов:

1) **Обсяг вибіркової сукупності.** Під час перевірки гіпотез за даними великих вибірок ( $n > 30$ ) доцільно застосовувати  $\chi^2$  -критерій нормального розподілу, а за даними малих вибірок ( $n < 30$ ) –t -статистику Стьюдента;

2) **Рівність вибірок за чисельністю.** Вибіркові сукупності за чисельністю можуть бути нерівними і рівними. Дані властивості необхідно враховувати під час практичної перевірки гіпотез про істотність відмінностей між середніми, зокрема під час розрахунку середньої помилки двох вибіркових середніх;

3) **Принципи формування вибірок.** Методи перевірки статистичних гіпотез залежать від характеру формування вибіркових сукупностей. Якщо спостереження однієї вибірки не протиставляються спостереженням іншої вибірки, то дані вибірки називають незалежними. Якщо ж спостереження однієї вибірки в деякій мірі пов'язані зі спостереженнями іншої вибірки, то такі вибірки називають залежними. Формування вибіркових сукупностей зумовлює різні прийоми оцінки вірогідності між середніми двох малих вибірок. Якщо вибірки незалежні, то статистичній оцінці підлягає різниця середніх, якщо залежні – середня різниця;

4) **Рівність дисперсій.** Під час перевірки гіпотез щодо середніх існує два випадки щодо вибіркових дисперсій: дисперсії рівні або нерівні. Таким чином виникає спеціальне завдання перевірки гіпотези про істотність відмінностей двох дисперсій. Для перевірки гіпотези про рівність двох дисперсій у генеральних сукупностях використовується критерій Фішера, який базується на основі співвідношень двох вибіркових скоригованих дисперсій, що замінюють значення дисперсій у генеральних сукупностях. Критичні показники критерію Фішера знаходять за спеціальними таблицями за відповідною кількістю ступенів свободи і по заданому рівню значущості.

Залежно від того, рівні чи нерівні дисперсії в генеральних сукупностях, виникає потреба у видозмінюванні схемі перевірки гіпотези. Для розв'язку задач другого типу котрі пов'язані з перевіркою гіпотез відносно законів розподілу генеральних сукупностей. У практичних дослідженнях часто виникає необхідність встановити характер розподілу генеральних сукупностей. В даному випадку можуть виникати наступні задачі:

- 1) про узгодженість емпіричного (вибіркового) і теоретичного (генерального) розподілу;
- 2) про незалежність розподілу однієї ознаки від другої;
- 3) про однорідність двох та більше емпіричних розподілів.

Зазначені гіпотези, такі, як ті, що стосуються параметрів розподілу, оцінюються за допомогою спеціальних критеріїв узгодження. Критерій узгодження - це метод оцінки гіпотези про очікуваний закон розподілу, який є невідомим для загальної сукупності. Ці критерії допомагають визначити, чи відповідають досліджувані розподіли теоретичним моделям розподілу, а також визначити ступінь значущості різниць між цими розподілами.

### **Основна й альтернативна гіпотези**

Одна гіпотеза називається основною (або нульовою), вона позначається  $H_0$ . В свою чергу, логічне та аргументоване заперечення основної гіпотези називається альтернативною та позначається  $H_1$ . Гіпотеза, що фіксує єдине значення параметра, називається простою. В свою чергу, гіпотеза, що документує декілька значень параметра, називається складною.

Статистичний критерій перевірки гіпотези  $H_0$  – це правило за яким приймається рішення про прийняття або відхилення гіпотези  $H_0$ . Рішення приймаються на основі вибірки  $X_1, X_2, \dots, X_n$ , за якою формують спеціальну

функцію вибірки  $T_n = T(X_1, X_2, \dots, X_n)$ , цю функцію називають статистикою критерію.

### Принцип перевірки гіпотез.

Множина можливих значень статистики критерію  $T_n$  розділяється на дві підмножини, що не перетинаються: критичну сферу  $S$ , тобто сферу відхилення гіпотези  $H_0$  та сферу  $S$  прийняття цієї гіпотези. Якщо фактичне (спостережене, вираховане на основі вибірки) значення критерію потрапляє в критичну сферу  $S$ , то основна гіпотеза  $H_0$  відхиляється та приймається альтернативна  $H_1$ , у протилежному випадку приймається основна гіпотеза, тобто відхиляється альтернативна. Як уже зазначалось раніше, під час перевірки статистичної гіпотези є певна ймовірність виникнення помилки.

Помилка першого роду полягає в тому, що відхиляється нульова гіпотеза, яка насправді є правильною. Ймовірність цієї помилки позначається  $\alpha$  та називається рівнем значущості критерію.

За означенням  $\alpha = P(H_1|H_0)$ . Цю ймовірність задають, як правило, перед проведенням тесту, числове значення є стандартним  $\alpha = 0,01; 0,05; 0,001; 0,005$ .

Помилка другого роду полягає у відхиленні альтернативної гіпотези, що насправді правильна. Ймовірність помилки другого роду позначається  $\beta$ . Відповідно до означення  $\beta = P(H_0|H_1)$ . Величина  $1 - \beta$  називається потужністю критерію. Мінімізувати ймовірність одночасно помилки 1-го та 2-го роду можливо лише за умови збільшення об'єму вибірки (табл. 3.1).

Таблиця 1.1

| Гіпотеза $H_0$     | Відхиляється                               | Приймається                                    |
|--------------------|--|--|
| <b>Правильна</b>   | Помилка 1-го роду<br>$\alpha = P(H_1 H_0)$ | Правильне рішення<br>$1 - \alpha = P(H_0 H_0)$ |
| <b>Неправильна</b> | Правильне рішення                          | Помилка 2-го роду                              |

|  |                          |                      |
|--|--------------------------|----------------------|
|  | $1 - \beta = P(H_1 H_1)$ | $\beta = P(H_0 H_1)$ |
|--|--------------------------|----------------------|

У радіолокації помилка першого роду – пропуск сигналу, другого роду – прийняття хибного сигналу; у юриспруденції  $\alpha$  – ймовірність виправдовування винного,  $\beta$  – засудження невинного; у виробництві  $\alpha$  – ризик постачальника,  $\beta$  – ризик споживача.

Рівень значимості гіпотези називають ймовірність здійснити похибку першого роду, тобто відхилити правильно нульову гіпотезу ( $\alpha = 10\%$  або  $5\%$  або  $1\%$ ). Площа під нормальним розподілом дорівнює 1. Нам необхідно перевірити, чи це випадкове середнє або середнє іншого розподілу. Похибка характеризує, що середнє значення попало випадково, тоді відхиляється нульова гіпотеза.

Розширюється площа, знижується відсоток похибки, на рівні  $0,05$  відхиляємо, а за  $0,01$  не відхиляємо (рис. 1.10).

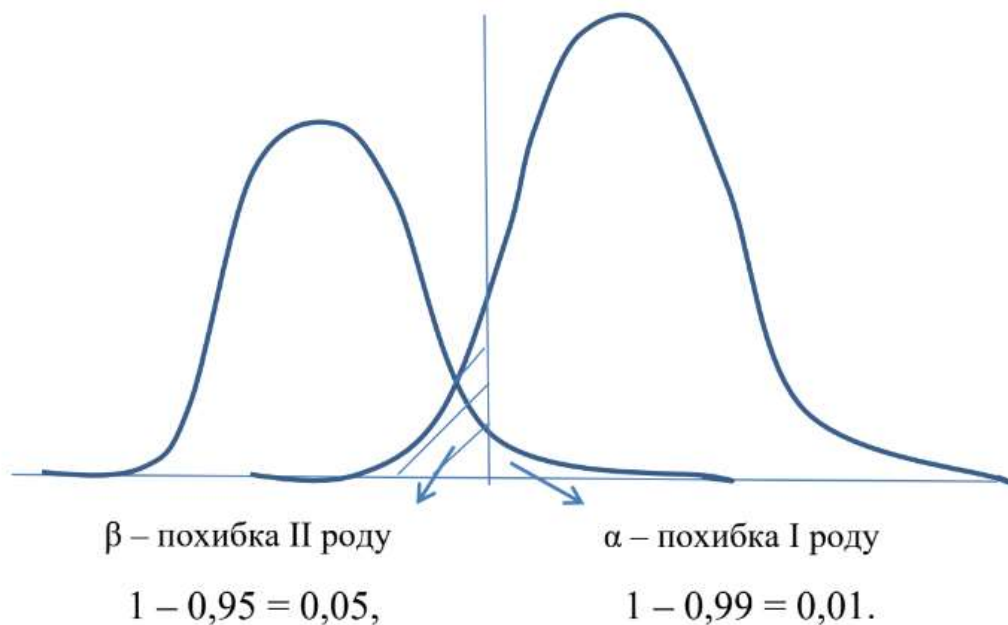


Рис 1.10. Візуалізація похибки першого і другого роду

Для візуалізації похибки застосуємо розподільчий калькулятор: [https://gallery.shinyapps.io/dist\\_calc/](https://gallery.shinyapps.io/dist_calc/) (рис. 1.11–1.12). Distribution Calculator (розподільчий калькулятор)

### Distribution Calculator

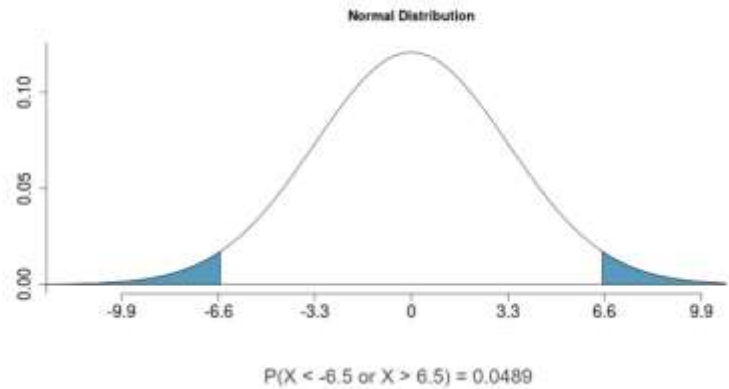


Рис 1.11 Нормальний розподіл

standard deviation – середньоквадратичне відхилення

mean – середнє значення

Find area – знайдіть сферу:

both tails – обидва хвости

lower tail – нижчий хвіст

upper tail – верхній хвіст

middle tails – середні хвости

### Distribution Calculator

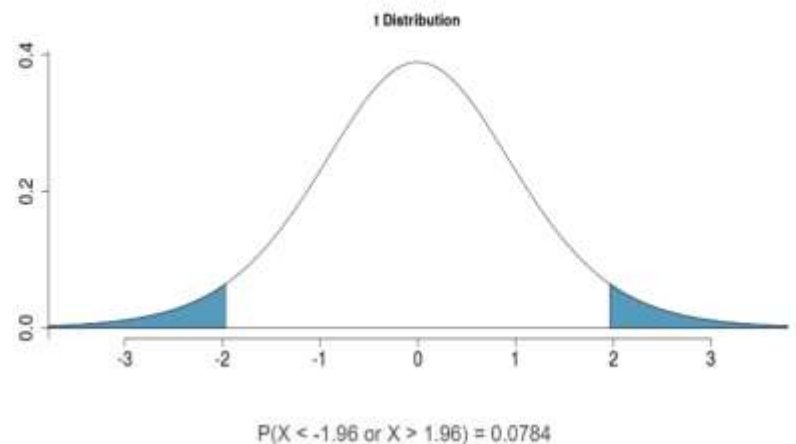


Рис. 1.12.  $t$ -розподіл

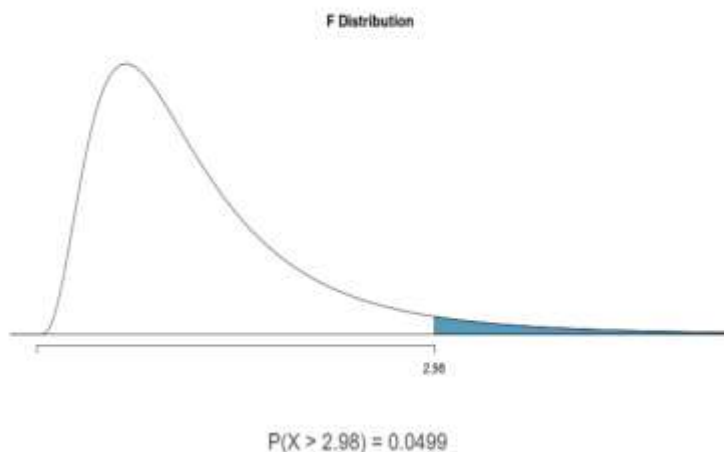


Рис.1.13.  $F$ -розподіл

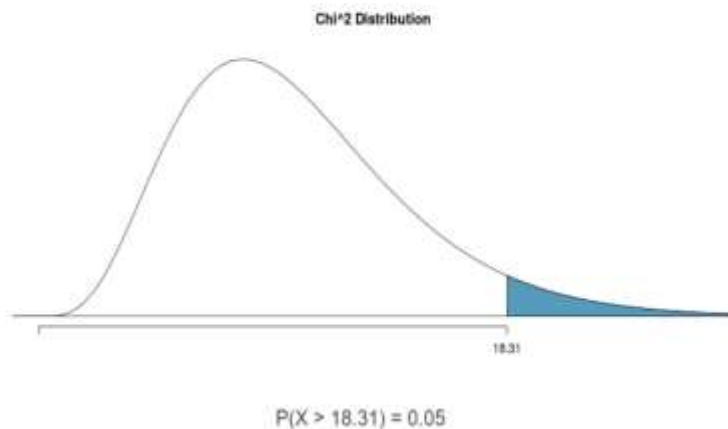


Рис.1.14.  $\chi^2$ -розподіл

### Алгоритм проведення статистичного тестування

Перевірка статистичної гіпотези передбачає послідовне виконання таких етапів:

1. Оцінити інформацію та описати статистичну модель вибіркової сукупності.



2. На основі вибірки  $X_1, X_2, \dots, X_n$  сформулювати нульову та альтернативну гіпотези.

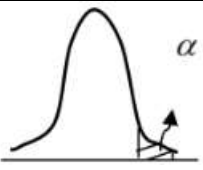
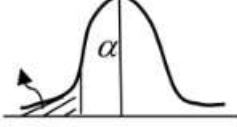
3. Вибрати критерій перевірки.

У кожному конкретному випадку підбирають статистику критерію  $T_n = T(X_1, X_2, \dots, X_n)$ , зазвичай з нижче перерахованих:  $U$  – нормальний розподіл, розподіл  $\chi^2$ -квадрат;  $t$ -розподіл Стюдента,  $F$ -розподіл Фішера.

4. За статистикою критерію та рівнем значущості  $\alpha$  у відповідній статистичній таблиці (у цьому підручнику ці таблиці подані в додатках) визначають критичне значення (критичну точку) та визначають критичну сферу  $S$ . Межі сфер визначаються, відповідно, із співвідношень для правосторонньої, лівосторонньої та двосторонньої сфер, що показано у таблиці 1.2.

Таблиця 1.2

Межі сфер статистичного критерію

| Критичні сфери         | Ймовірність помилки 1-го роду                          | Візуалізація  |
|------------------------|--|---|
| Правостороння          | $P(t_{сп} > t_{кр}) = \alpha$                          |  |
| Лівостороння           | $P(t_{сп} < t_{кр}) = \alpha$                          |  |
| Двостороння            | $P(t_{сп} < t_{кр1}) + P(t_{сп} > t_{кр2}) = \alpha$   |  |
| Симетрична двостороння | $P(t_{сп} < t_{кр1}) = P(t_{сп} > t_{кр2}) = \alpha/2$ |   |

5. Для отриманої реалізації вибірки  $x_1, x_2, \dots, x_n$  розраховують фактичне значення критерію  $T_{спостер} = T(x_1, x_2, \dots, x_n) = t$ .

6. Якщо розраховане значення  $t$  належить критичній сфері, то гіпотезу  $H_0$  відхиляють; якщо ж  $t \notin S$ , тоді немає підстав для відхилення основної гіпотези, тобто  $H_0$  приймають.

7. Формулювання висновків за результатами перевірки нульової гіпотези.

Необхідно зауважити, що у разі відсутності статистичних таблиць із критичним значенням можна скористатися значимістю критерію (р-значення).

Приклад. Побудувати гістограму за даними вибірки та зробити припущення про закон розподілу.

Таблиця 1.2

Вихідні дані

| $x_i; x_{i+1}$ | 0-4 | 4-8 | 8-12 | 12-16 | 16-20 | 20-24 |
|----------------|-----|-----|------|-------|-------|-------|
| $n_j$          | 40  | 24  | 16   | 12    | 8     | 4     |
| $n_j/h$        | 10  | 6   | 4    | 3     | 2     | 1     |

$$h=4-0=4.$$

Сполучивши середини прямокутників гістограми, за виглядом отриманої кривої можемо зробити припущення, тобто висунути гіпотезу про існування експоненціального розподілу. Кожну гіпотезу варто перевірити, а для цього потрібні емпіричні та теоретичні частоти. Емпіричні – це частоти знайдені за даними вибірки, теоретичні частоти – обчислюємо за формулами.

У випадку дискретного закону розподілу теоретичні частоти шукають за формулою:  $n'_i = n \cdot p_i$ , де  $n$  – об'єм вибірки;  $p_i$  – ймовірність спостережуваного значення  $X = x_i$ , яка обчислюється заумови, що  $X$  взяте за припущеним законом.

У більшості випадків закон розподілу досліджуваної випадкової величини невідомий, але є певні підстави для припущення, що він є, наприклад, нормальним, показниковим, рівномірним, гіпергеометричним тощо.

Нехай необхідно перевірити гіпотезу  $H_0$  про те, що випадкова величина розподілена за певним законом, що задається функцією розподілу  $F_0(x)$ , тобто основна гіпотеза має вигляд  $H_0: F_X(x) = F_0(x)$ .

Альтернативна гіпотеза суперечить основній, тобто  $H_1: F_X(x) \neq F_0(x)$ . Для перевірки гіпотези про розподіл випадкової величини  $X$  проводимо вибірку і представляємо її у формі таблиці 1.3 статистичного розподілу.

Таблиця 1.3

Вибірка статистичного розподілу

|       |       |       |     |       |
|-------|-------|-------|-----|-------|
| $x_i$ | $x_1$ | $x_2$ | ... | $x_m$ |
| $n_j$ | $n_1$ | $n_2$ | ... | $n_m$ |

З об'ємом вибірки  $n = \sum_{i=1}^m n_i$

### Види розподілів

Дискретна випадкова величина  $X$  має біноміальний закон розподілу з параметрами  $n \in N, p \in [0; 1]$ , якщо вона набуває значення  $0, 1, 2, \dots, m, \dots, n$  з імовірностями, що обчислюються за формулою Бернуллі:  $P(X = m) = C_n^m p^m q^{n-m}$

Той факт, що випадкова величина  $X$ , розподілена за біноміальним законом з параметрами позначається таким чином  $X \sim B(n; p)$  (табл. 3.5).

Таблиця 1.4

Вибірка біноміального розподілу

|       |       |                     |                     |     |                     |     |       |
|-------|-------|---------------------|---------------------|-----|---------------------|-----|-------|
| $x_i$ | 0     | 1                   | 2                   | ... | M                   | ... | n     |
| $p_j$ | $q^n$ | $C_n^1 p^1 q^{n-1}$ | $C_n^2 p^2 q^{n-2}$ | ... | $C_n^m p^m q^{n-m}$ | ... | $p_n$ |

Біноміальний закон розподілу широко використовується в теорії і практиці статистичного контролю якості продукції, під час описування функціонування систем масового обслуговування та в інших сферах.

Математичне сподівання випадкової величини  $X$ , розподіленої за біноміальним законом, обчислюється за формулою  $M(X) = np$ .

Дисперсія випадкової величини  $X$ , розподіленої за біноміальним законом, обчислюється за формулою  $D(X) = npq$ .

Дискретна випадкова величина  $X$  має розподіл Пуассона з параметром  $\lambda = np$ ,  $0 < \lambda < \infty$ ,  $n \in N$ ,  $p \in (0; 1)$ , якщо вона приймає значення  $0, 1, 2, \dots, m, \dots$  – це кількість успіхів у  $n$  незалежних випробуваннях Бернуллі, за умови якщо  $n \rightarrow +\infty$ ,  $p \rightarrow 0$  з ймовірностями, що шукають за формулою (табл. 3.6):  $P(X = m) = \frac{\lambda^m e^{-\lambda}}{m!}$

Таблиця 1.5

Вибірка розподілу Пуассона

|       |                |                        |                                     |     |                                     |     |
|-------|----------------|------------------------|-------------------------------------|-----|-------------------------------------|-----|
| $x_i$ | 0              | 1                      | 2                                   | ... | $m$                                 | ... |
| $p_i$ | $e^{-\lambda}$ | $\lambda e^{-\lambda}$ | $\frac{\lambda^2}{2!} e^{-\lambda}$ | ... | $\frac{\lambda^m e^{-\lambda}}{m!}$ | ... |

Розподіл Пуассона за параметром  $\lambda$  позначається таким чином:  $PS(\lambda)$ .

За допомогою цього закону досліджується кількість збоїв на автоматичній лінії, відмов складної системи, що працює в нормальному режимі, вимог на обслуговування, які поступають за одиницю часу в системах масового обслуговування тощо.

Якщо випадкова величина  $X$  має розподіл Пуассона з параметром, тоді: математичне сподівання  $M(X) = \lambda$ ; дисперсія  $D(X) = \lambda$ .

## **Висновки до 1 розділу**

В даному розділі ми розглянули базові поняття бізнес-аналітики багатовимірних процесів. Також була детально розкрита тема сутності та можливих методів застосувань багатовимірного статистичного аналізу в бізнес-аналітиці. В контексті особливостей обробки багатовимірних статистичних даних, було розглянуто види просторових ознак та етапи досліджень за допомогою статистичного аналізу. Також, ми розглянули принципи перевірки статичних гіпотез, та алгоритми проведення статистичного тестування.

## **РОЗДІЛ 2. ОСОБЛИВОСТІ СУЧАСНОГО ПРОГРАМНОГО ЗАБЕЗПЕЧЕННЯ РОЗВ'ЯЗАННЯ ЗАДАЧА БАГАТОВИМІРНОГО СТАТИСТИЧНОГО АНАЛІЗУ**

### **2.1. Аналіз сучасного стану на ринку програмного забезпечення статистичної обробки даних**

Аналіз даних являється невід'ємною складовою процесу дослідження систем, котрі відносяться до будь-якого типу. В наші дні, ринок програмного забезпечення (ПЗ) для статистичної обробки даних представлений широкою лінійкою різноманітного ПЗ. У даному списку запропоновані різноманітні ПЗ від засобів EXCEL до пакету STATISTICA.

При здійсненні науково-дослідницької роботи важливо ретельно обирати статистичний пакет, відповідно до завдань, які постають перед користувачем. Зазвичай, найбільш оптимальним вибором є програмне забезпечення, яке поєднує високу продуктивність, необхідний функціонал і доступну ціну.

При виборі слід приділити увагу наступним аспектам: перевірити, чи задовольняє комп'ютерне обладнання користувача системним вимогам програмного забезпечення; переконатися, що обране програмне забезпечення вміє виконувати необхідні завдання; оцінити, чи розраховане ПЗ на очікуваний обсяг даних для статистичного аналізу; врахувати рівень знань користувача в галузі статистики.

Статистичний пакет повинен мати такі важливі характеристики:

- модульність для вибору потрібних функцій;
- можливість надавати підказки та поради користувачу при виборі методів обробки даних;

- використання легкої та зрозумілої мови для постановки завдань; автоматизований процес обробки даних та взаємозв'язок між різними модулями ПЗ;
- збереження користувацьких даних та можливість генерації звітів;
- можливість взаємодії в діалоговому режимі;
- сумісність із іншими програмами.

Для використання кожного ПЗ наявна інформаційна система повинна задовільняти відповідним системним вимогам. Системні вимоги (System Requirements) – це мінімальні системні вимоги що фіксують, якою має бути конфігурація системи необхідною для запуску програмного забезпечення або так звані рекомендовані вимоги до системи, що показують, яка конфігурація системи здатна забезпечити максимально комфортні та прийнятні умови роботи ПЗ. Ці вимоги можуть описувати, як програмне забезпечення (software - необхідні драйвери, операційна система тощо), так і апаратне забезпечення (hardware). Для прикладу, основні системні вимоги можуть бути такими:

- частота роботи процесора - 2,5 ГГц;
- операційна система - Windows 10 (32- та 64-розрядна версія) або Linux Debian 12 (BSD, GPL);
- розмір оперативної пам'яті - 1 ГБайт;
- розмір вільного простору на жорсткому диску - 24,5 ГБайт;
- наявність браузера – Google Chrome 8, 9, 10, 11 або Firefox (ESR);
- роздільна здатність екрана - 1024x768 пікселів;
- необхідна відео карта - ATI Radeon 9400/128 Мбайт (або апаратне прискорення відео додатково);
- необхідна DirectX-сумісна аудіокарта.

Практика використання комп'ютерної техніки та програмного забезпечення доводить, що для майже кожної науково-дослідницької задачі існує можливість знаходити оптимальну комбінацію апаратних засобів та програмних рішень. Існують різні варіанти програмного забезпечення та різні версії цього ПЗ з різними системними вимогами, які можна використовувати на різних апаратних платформах. Значущим є той факт, що розвиток як програмного, так і апаратного забезпечення відбувається паралельно і невинно. Отже, обрати підходящу систему інструментів для власних досліджень не повинно бути проблемою.

Для досягнення оптимальної продуктивності важливі такі аспекти програмного забезпечення: можливість розширення функціоналу, високий рівень сегментації, наявність різних методик аналізу та сервісних функцій, розширені засоби налаштувань, зручні можливості імпорту даних та адаптація до специфіки конкретної галузі.

## **2.2. Дослідження характеристик програмних засобів для математичної і статистичної обробки даних**

На ринку статистичних пакетів існує велика кількість продуктів, яка налічує декілька тисяч. За функціональністю, статистичні програми можна розділити на три основні категорії: універсальні пакети, професійні пакети і спеціалізовані пакети. Наприклад, до професійних пакетів належать SAS і BMDP, універсальними є STADIA, STATGRAPHICS, SPSS та STATISTICA, а спеціалізовані призначені для вузьких областей і включають BIOSTAT, MESOSAUR, DATASCOPE.

Статистичні програми вважаються складними та наукомісткими програмами, і їх вартість часто перевищує можливості індивідуальних користувачів. Професійні пакети надають широкий спектр методів аналізу, тоді як популярні універсальні пакети мають достатньо функцій для загального



застосування. Спеціалізовані пакети орієнтовані на конкретну область аналізу даних. Розробники програм статистичних пакетів стверджують, що їх продукт перевершує конкурентів. Проте в більшості дослідників відсутній час на освоєння декількох програм, що ускладнює вибір необхідного програмного забезпечення.

Інформаційні системи активно використовують дані різних типів, і для аналізу цих даних існують класичні методи, що ґрунтуються на математичних концепціях, таких як математична статистика, математичне програмування, лінійна алгебра та інші. Ці методи мали успішне застосування протягом багатьох років.

Сучасне програмне забезпечення дозволяє здійснювати повноцінний математичний та статистичний аналіз даних. На ринку програмного забезпечення існує велика кількість додатків, які надають допомогу в аналізі даних, як у пакетному режимі, так і у вигляді бібліотек функцій, які можуть бути використані в інших програмних продуктах.

Більшість класичних пакетів, таких як RATS, SAS, STATA, TSP, SPSS, SORITEC, SYSTAT, були створені до 80-х років і орієнтовані на великі ЕОМ (електронно-обчислювальні машини). Фактично вони були колекцією процедур, написаних на мові FORTRAN. Багато з цих пакетів досі включають частини, написані на FORTRAN. Під час перенесення на персональні комп'ютери над цими процедурами додавалася оболонка у вигляді командної мови для управління процедурами (або досить простої системи меню, яка базується на командній мові). У зв'язку з цим вони не такі продуктивні, як спеціалізовані пакети, призначені для роботи на ПК (наприклад, GAUSS, Microfit, MicroTSP, EViews). Ці спеціалізовані пакети вигідно виділяються за такими характеристиками, як швидкодія, дизайн, інтерфейс користувача та ефективне використання ПК.

Програмні пакети, розроблені для персональних комп'ютерів, зазвичай написані на мові програмування C, що дозволяє створювати більш гнучкий та ефективний інтерфейс для користувача. Однак, пакети, що спочатку були реалізовані на мові FORTRAN, такі як TSP, STATA, SPSS, SYSTAT, були розроблені і вдосконалювалися великими групами розробників протягом багатьох років, і вони включають в себе значну кількість різноманітних методів та функцій.

Під час вибору програмного забезпечення для статистичного аналізу даних, важливо враховувати можливість легкості створення графіків, включаючи отримання графіків безпосередньо з меню аналізу регресії, а також наявність інтерактивних графічних можливостей, таких як графічний курсор і графічний редактор.

Тепер розглянемо деякі з найпопулярніших та функціонально повних програмних продуктів для статистичного аналізу даних.

**BMDP**, або BioMeDical Package, розроблений компанією Statistical Software, Inc., представляє собою біомедичний статистичний пакет і бібліотеку, яка включає понад сорок різних статистичних процедур. Перша версія BMDP була написана на мові Fortran для комп'ютера IBM 7090. BMDP є еволюційним розвитком вихідних BIMED програм та була розроблена в Університеті Каліфорнії в Лос-Анджелесі (UCLA) в галузі медичних наук.

Серед функцій, які надає BMDP, можна виділити такі:

- Ступенева регресія
- Визначення всіх можливих підмножин регресії
- Поліноміальна регресія
- Ступенева логістична регресія
- Дисперсійний та коваріаційний аналіз

ВМДР є потужним інструментом для аналізу даних у галузі біомедицини та інших медичних наук.

**EViews**, що означає Econometric Views, розроблений компанією Quantitative Micro Software (зараз власність IHS Markit Ltd.), є статистичним пакетом, спеціалізованим на аналізі економетричних даних та часових рядів. Цей програмний продукт дозволяє проводити аналіз, моделювання та побудову регресійних моделей, а також вирішувати різноманітні завдання в галузі економіки та фінансів.

EViews відомий своєю потужністю і можливістю вирішувати такі завдання:

- Аналіз наукової інформації
- Фінансовий аналіз
- Макроекономічне прогнозування
- Моделювання економічних процесів
- Прогнозування стану ринків та інші.

Особливі можливості EViews виявляються під час аналізу часових рядів. Деякі з його функцій включають:

- Проста лінійна регресія
- Багатофакторна регресія
- Нелінійна регресія
- Симультаивні моделі
- ARIMA-моделі

Один з переваг EViews полягає в тому, що він має зручний і логічний інтерфейс для користувача, а також обширну систему довідки, що допомагає засвоїти економетричні методи та підходи.

**GenStat**, розроблений компанією VSN International Ltd. (VSNi), є програмним пакетом для статистичного аналізу даних. Ця програма володіє розширеними можливостями візуалізації результатів та має вбудовану систему програмування. Важливою особливістю GenStat є наявність передбачених функцій для аналізу результатів біологічних, фармакологічних і сільськогосподарських експериментів, включаючи аналіз мікроматриць (microarray).

GenStat є вельми різностороннім інструментом статистичного аналізу, що пропонує:

- Простий інтерфейс через меню Windows для новачків та користувачів з обмеженим досвідом.
- Потужний інтерфейс з використанням командної мови для більш досвідчених користувачів.

Основною перевагою GenStat є великий спектр доступних статистичних методів, що дозволяє використовувати його для різних дослідницьких завдань.

**GraphPad Prism**, розроблений компанією GraphPad Software, Inc., є спеціалізованою програмою для статистичного аналізу біологічних даних, зокрема біостатистики та побудови графіків та кривих. Це високоефективний інструмент, розроблений для дослідників у галузі біології, соціології, фізики та інших наукових галузей.

GraphPad Prism володіє такими ключовими функціями:

1. Побудова наукових графіків: Програма дозволяє побудовувати графіки для візуалізації даних, що допомагає зрозуміти зв'язки та тенденції в дослідженні.
2. Нелінійна регресія: GraphPad Prism спрощує аналіз нелінійних залежностей, дозволяючи обирати зі списку рівнянь та автоматично побудовувати криві та робити інтерполяцію невідомих значень.

3. Статистичний аналіз: Програма надає широкий спектр статистичних тестів, таких як t-тест, непараметричні порівняння, дисперсійний аналіз, аналіз таблиць зв'язаності та виживання, для оцінки результатів дослідження.
4. Легкість використання: GraphPad Prism розроблений так, щоб бути зручним як для досвідчених дослідників, так і для початківців. Інтерфейс зрозумілий та дозволяє легко працювати з даними.
5. Пов'язаність даних: Всі частини проекту в програмі пов'язані, що дозволяє внесені зміни в даних автоматично впливати на результати і графіки.

GraphPad Prism є корисним інструментом для дослідників та студентів, які працюють з біологічними даними і вимагають зручного і потужного пакету для статистичного аналізу та візуалізації даних.

**Maple**, розроблений компанією Waterloo Maple, Inc., є комерційною системою комп'ютерної алгебри і забезпечує велику кількість функцій для різних областей математики та наукових досліджень. Основні можливості цього програмного продукту включають:

1. Символьні та чисельні обчислення: Maple надає можливість виконувати обчислення як символьно, так і чисельно, що робить його корисним інструментом для різних математичних завдань.
2. Математичні функції та методи: Програма містить широкий набір математичних функцій і методів для виконання різних математичних операцій.
3. Розв'язування рівнянь і диференціальних рівнянь: Maple допомагає розв'язувати рівняння та диференціальні рівняння, що корисно для фізичних наук, інженерії та інших галузей.
4. Лінійна алгебра та оптимізація: Програма підтримує операції з лінійною алгеброю і методи оптимізації, що важливо для різних наукових досліджень.

5. Програмування: Maple має власну мову програмування, що дозволяє користувачам створювати власні алгоритми та програми для рішення конкретних завдань.
6. Візуалізація та графіка: Програма дозволяє побудовувати графіки і візуалізувати дані, що сприяє кращому розумінню результатів.
7. Інтерактивність: Maple має інтерактивний інтерфейс, що сприяє зручності роботи з даними і аналізом результатів.
8. Моделювання: Ви можете використовувати Maple для моделювання різних явищ і процесів, включаючи фізичні моделі та статистичне моделювання.
9. Підтримка CUDA і багатопроцесорних обчислень: Maple підтримує використання графічних процесорів (CUDA) і багатопроцесорних систем для високопродуктивних обчислень.
10. Інтеграція з іншими мовами програмування: Програма дозволяє взаємодіяти з іншими мовами програмування, включаючи MATLAB.
11. Зв'язок з базами даних: Maple має можливості підключення до баз даних для обробки та аналізу даних.

Завдяки широкому спектру функцій, Maple є потужним інструментом для вчених, інженерів та дослідників у багатьох наукових галузях.

**MathCAD** - це потужний пакет для математичних розрахунків, що дозволяє виконувати різноманітні обчислення, вирішувати рівняння та будувати графіки. Він належить до класу систем автоматизованого проектування і призначений для створення інтерактивних документів з обчисленнями і візуальним супроводженням. Основні особливості MathCAD:

1. Комп'ютерна алгебра та обчислення: MathCAD - це система, що використовує комп'ютерну алгебру для розв'язання математичних задач та обчислень.

2. Інтуїтивний інтерфейс: Програма має простий та інтуїтивний інтерфейс, що дозволяє легко працювати з формулами та даними за допомогою клавіатури або спеціальних панелей інструментів.
3. Візуалізація результатів: Робота ведеться на робочому аркуші, де рівняння та вирази відображаються графічно, що сприяє кращому розумінню та візуалізації математичних виразів.
4. Принцип WYSIWYG: Методологія "What You See Is What You Get" дозволяє користувачам отримувати результати обчислень такими, якими вони бачать на екрані.
5. Використання в навчанні та складних проектах: Ця програма підходить як для навчання, так і для складніших проектів, що вимагають візуалізації результатів математичного моделювання.
6. Інтеграція з іншими технологіями: Завдяки відкритій архітектурі та підтримці технологій NET і XML, MathCAD легко інтегрується з іншими ІТ-структурами та інженерними застосуваннями.
7. Створення електронних книг: Можливість створювати електронні книги (e-Book), що дозволяє зберігати та поширювати матеріали в зручній електронній формі.

MathCAD є зручним інструментом для проведення різних обчислень, інженерних розрахунків та для створення візуальних математичних документів, що корисно для вчителів, студентів, науковців і інженерів.

**Mathematica**, розроблена компанією Wolfram Research, є потужною системою комп'ютерної алгебри, яка володіє широким спектром можливостей для чисельних, символічних та графічних обчислень і візуалізації. Ось деякі ключові особливості та можливості Mathematica:

1. Символьні обчислення: Mathematica може виконувати символічні обчислення, що дозволяють вирішувати різні математичні завдання в символічному вигляді. Вона може працювати зі змінними, функціями, поліномами, матрицями та іншими алгебраїчними виразами.
2. Чисельні обчислення: Mathematica також підтримує чисельні обчислення з різних областей, включаючи чисельний аналіз, диференціальні рівняння, оптимізацію і чисельне інтегрування.
3. Графіка: Програма дозволяє легко побудовувати графіки функцій, кривих, поверхонь та інших графічних об'єктів.
4. Мова програмування: В Mathematica є вбудована мова програмування, яка дозволяє користувачам створювати власні програми та розв'язувати складні задачі.
5. Робота зі звітами і документами: Mathematica дозволяє створювати "записники" (notebooks), що об'єднують текст, обчислення та графіку в одному документі. Це зручно для створення технічних звітів і презентацій.
6. Пакети додатків: Існує можливість розширення функціональності Mathematica завдяки пакетам додатків, які можна легко встановити та використовувати.

Mathematica користується популярністю в багатьох галузях, включаючи математику, фізику, інженерію, комп'ютерні науки, фінанси та інші. Вона дозволяє вченим і професіоналам здійснювати складні математичні обчислення та аналіз, а також вирішувати різноманітні завдання у своїх галузях

Mathematica дійсно пропонує широкі можливості для аналітичних та чисельних обчислень у різних математичних галузях. Давайте розглянемо більше деталей щодо цих можливостей:

#### Аналітичні перетворення:

1. Розв'язання рівнянь: Mathematica може розв'язувати різні типи рівнянь, включаючи поліноміальні, тригонометричні та трансцендентні рівняння.



2. Рекурентні рівняння: Вона може розв'язувати рекурентні рівняння, що зустрічаються в численних областях.
3. Спрощення виразів: Mathematica може автоматично спрощувати складні математичні вирази.
4. Знаходження границь: Програма дозволяє обчислювати границі функцій.
5. Інтегрування і диференціювання: Вона розв'язує інтегральні та диференціальні рівняння, обчислює похідні та інтеграли.
6. Робота зі скінченними і нескінченними сумами і добутками: Mathematica допомагає у вирішенні завдань, пов'язаних із сумами і добутками.
7. Перетворення Фур'є, Лапласа та Z-перетворення: Для аналізу сигналів і сигнальних систем.
8. Ряди Тейлора: Mathematica може знаходити ряди Тейлора функцій та виконувати операції з ними, такі як додавання, множення, композиція та отримання зворотної функції.

#### Чисельні розрахунки:

1. Обчислення значень функцій: Mathematica дозволяє обчислювати значення функцій, включаючи спеціальні функції, з великою точністю.
2. Рішення систем рівнянь: Можливість розв'язувати системи лінійних та нелінійних рівнянь.
3. Інтегрування і диференціювання: Для обчислення інтегралів та похідних функцій.
4. Поліноміальна інтерполяція: Для наближення функцій за допомогою поліномів.

#### Завдання у теорії чисел:

1. Визначення простих чисел: Mathematica може допомогти визначити прості числа за їх порядковим номером.

2. Розкладання чисел на прості множники: Допомагає знайти розклад числа на прості множники.

3. Дискретне перетворення Фур'є: Важлива операція для аналізу сигналів і даних у цифровому вигляді.

Лінійна алгебра:

1. Операції з матрицями: Mathematica дозволяє виконувати операції з матрицями, включаючи додавання, множення, знаходження оберненої матриці, обчислення експоненти, знаходження власних значень і власних векторів.

Графіка і звук:

2. Побудова графіків: Mathematica може побудовувати графіки функцій, параметричних кривих та поверхонь.

3. Побудова геометричних фігур: Допомагає побудувати геометричні об'єкти, такі як ламані, кола та прямокутники.

4. Робота зі звуком: Mathematica дозволяє створювати і обробляти звукові сигнали та супроводжувати їх аналітичною функцією або набором точок.

5. Імпорт і експорт графіки та звуку: Можливість імпорту та експорту даних у різних форматах для обробки та аналізу.

Mathematica – потужний інструмент для математичних обчислень та аналізу, який знаходить застосування в багатьох галузях науки і техніки.

**MATLAB** (Matrix Laboratory) - це потужна інтегрована система обчислень та візуалізації, розроблена компанією MathWorks, Inc. Вона широко використовується для вирішення різноманітних математичних, інженерних і наукових завдань. Основні характеристики і можливості MATLAB включають:

1. Мова програмування MATLAB: MATLAB має власну мову програмування, яка включає в себе обробку масивів та матриць, структури даних, оператори

контролю потоку та велику кількість вбудованих функцій для чисельних обчислень.

2. Робота з матрицями та векторами: MATLAB відомий своєю потужною системою роботи з матрицями. Він дозволяє виконувати операції з матрицями, лінійну алгебру, розв'язувати лінійні рівняння та визначати власні значення.

3. Чисельний аналіз: MATLAB надає широкий спектр вбудованих функцій для чисельного аналізу, включаючи розв'язання диференціальних рівнянь, оптимізацію, інтегрування та чисельні методи.

4. Графічна візуалізація: MATLAB надає засоби для побудови графіків та візуалізації даних. Ви можете створювати 2D та 3D графіки, діаграми, анімації та інтерактивні візуалізації.

5. Підтримка додатків: Ви можете створювати власні додатки та інтерфейси користувача в MATLAB, що робить його корисним для розробки власних програм та інтерактивних додатків.

6. Інтеграція з іншими мовами та програмами: MATLAB може бути інтегрований з іншими мовами програмування, такими як C, C++, Java, і взаємодіяти з іншими програмами та обчислювальними середовищами.

7. Застосування в науці та промисловості: MATLAB широко використовується в різних галузях, включаючи інженерію, фізику, економіку, обробку сигналів, машинне навчання та інші області.

Це робить MATLAB потужним інструментом для вирішення різноманітних завдань в чисельних обчисленнях, дослідженнях та розробці програм.

**MedCalc** - це спеціалізоване програмне забезпечення для статистичного аналізу даних, яке зосереджене на обробці і аналізі медичних та біологічних даних. Ось деякі з його можливостей:

1. ROC-аналіз (Receiver Operating Characteristic): MedCalc спеціалізується на аналізі ROC-кривих, які використовуються для оцінки діагностичних тестів та моделей класифікації в медичній статистиці. Він дозволяє побудовувати та аналізувати ці криві для оцінки точності та ефективності діагностичних тестів.

2. Обчислення статистичних параметрів: MedCalc підтримує обчислення різних статистичних параметрів, таких як середнє значення, стандартне відхилення, довірчі інтервали тощо.

3. Аналіз даних для медичних експериментів: Програма має можливості для аналізу даних, які часто зустрічаються в медичних експериментах, включаючи тестування гіпотез, аналіз залежності та кореляції між змінними.

4. Побудова графіків: MedCalc дозволяє побудовувати різноманітні графіки для візуалізації даних, результатів тестів та статистичного аналізу.

5. Медична статистика: Програма має вбудовані статистичні методи, які використовуються для медичних досліджень, такі як аналіз чутливості та специфічності, розрахунок відносного ризику, числові характеристики для діагностичних тестів тощо.

6. Довірчі інтервали та р-значення: Програма автоматично розраховує довірчі інтервали та р-значення для статистичних аналізів.

MedCalc спеціалізується на конкретних завданнях, пов'язаних із медициною та біологією, і надає інструменти, які часто вимагаються для обробки та аналізу даних в цих галузях. Програма відома своєю зручністю та простотою, що дозволяє легко виконувати аналіз та отримувати результати в медичному дослідженні.

**MESOSAUR** – МЕЗОЗАВР (СП «Стат-Диалог») - це програмне забезпечення для аналізу часових рядів, яке розроблене для персональних комп'ютерів IBM PC/XT/AT. Воно призначене для використання фахівцями з середньою кваліфікацією, які мають певне розуміння методів аналізу часових рядів. Основні можливості МЕЗОЗАВР включають:

1. Ділова графіка: Програма дозволяє візуалізувати дані у вигляді графіків, що може допомогти в зрозумінні патернів і тенденцій в часових рядах.
2. Параметричні тести: МЕЗОЗАВР має вбудовані параметричні статистичні тести та функції для описової статистики.
3. Непараметричні тести: Включають хі-квадрат, біноміальний тест, тести знаків, тести на послідовності, а також обчислення кореляції Кенделла та кореляції Спірмана.
4. Регресійний аналіз: Можливості включають просту регресію, множинну регресію, покрокову регресію, робасту регресію та нелінійну регресію.
5. Аналіз часових рядів: Включає функції згладжування, фільтрації, автокореляції, кроскореляції, аналізу спектру та кроспектрального аналізу.
6. Модель Бокса-Дженкінса (ARIMA): Програма підтримує аналіз та моделювання часових рядів за допомогою моделі ARIMA з підтримкою експертного аналізу.
7. Часовий контекст: МЕЗОЗАВР автоматично зіставляє відповідний момент часу для кожного значення часового ряду, що спрощує аналіз та побудову моделей.
8. Зберігання та імпорт даних: Програма має власний стандарт файлів і дозволяє зручно вводити та зберігати дані через вбудований редактор даних типу «електронної таблиці». Також підтримується експорт та імпорт даних з текстових файлів.

МЕЗОЗАВР призначений для обробки часових рядів з помірною кількістю спостережень, і він допомагає виконувати різні види статистичного аналізу, які є корисними для багатьох галузей, включаючи економіку, демографію, техніку, медицину тощо.

Дані параметри стосуються обмежень на розмір та кількість значень часових рядів в програмі МЕЗОЗАВР. Важливо розуміти ці обмеження при використанні програми. Основні обмеження включають:

1. Довжина часового ряду: Гранична довжина аналізованого часового ряду дорівнює 16 тисяч значень. Зазвичай програма найефективніше працює з рядами до 2-3 тисяч значень. Тобто, якщо ваш часовий ряд має більше 16 тисяч значень, може виникнути обмеження.

2. Кількість змінних: Одночасно можна аналізувати до 256 різних змінних. Проте сумарна кількість значень не повинна перевищувати 60 тисяч чисел.

3. Часова шкала: Ви можете працювати з реальними часовими шкалами, де крок за часом може бути від 1 хвилини до будь-якого цілого числа років. Або ж ви можете використовувати умовну часову шкалу.

4. Редагування даних: Для редагування даних в програмі МЕЗОЗАВР є вбудований табличний редактор, а також графічний режим для внесення змін.

5. Перетворення даних: Можливості програми включають перетворення даних за допомогою спеціального меню перетворень. Важливо вказати попереднє перетворення даних, яке враховується в процесі аналізу.

Ці обмеження важливі для того, щоб належним чином використовувати програму МЕЗОЗАВР і розраховувати на правильні результати аналізу часових рядів. При перевищенні цих обмежень можуть виникнути проблеми з продуктивністю програми або недостатньою точністю результатів.

**MicroTSP**, або Time Series Processor, є програмою для аналізу часових рядів та прогнозування. Вона надає різні функції для обробки та моделювання часових рядів. Основні функції пакету включають:

1. Проста регресія: Можливість проводити аналіз регресії для з'ясування взаємозв'язку між двома або більше змінними.

2. Регресія часових рядів: Використання регресійних моделей для аналізу та прогнозування часових рядів.

3. Прогнозування: Можливість створювати прогнози на основі аналізу часових рядів.

4. Авторегресивні моделі та моделі ковзного середнього: Використання авторегресивних моделей (AR), моделей ковзного середнього (MA) та їх комбінацій (ARMA) для моделювання та прогнозування часових рядів.

5. Одночасні (симультаивні) та імітаційні моделі: Можливість створювати моделі, які враховують одночасні взаємозв'язки між змінними. Імітаційні моделі дозволяють створювати сценарії та симулювати події в часових рядах.

6. Використання електронних таблиць: Взаємодія із зовнішніми електронними таблицями у поєднанні з MicroTSP для аналізу та обробки даних.

7. Логістичні та ймовірнісні моделі: Можливість використовувати логістичні та ймовірнісні моделі для аналізу часових рядів і прогнозування.

Ці функції роблять MicroTSP потужним інструментом для аналізу і обробки часових рядів у різних галузях, включаючи фінанси, економіку, науку та інженерію.

**Minitab** - це програмний пакет для статистичного аналізу та обробки даних, який широко використовується в наукових дослідженнях та промисловості. Він надає широкий спектр функцій для аналізу даних. Основні функції Minitab включають:

1. Управління даними і файлами: Мінітаб дозволяє імпортувати, організувати та аналізувати дані з різних джерел.

2. Електронні таблиці: Minitab надає зручний інтерфейс для введення та організації даних у вигляді електронних таблиць, що полегшує їх аналіз.

3. Регресійний аналіз: Включає можливість проводити регресійний аналіз, включаючи просту регресію та регресію часових рядів.

4. Потужність і розмір вибірки: Допомагає визначити оптимальний розмір вибірки для проведення статистичних досліджень.
5. Таблиці і діаграми: Дозволяє створювати різні таблиці та діаграми для візуалізації даних.
6. Багатовимірний аналіз: Включає факторний аналіз, кластерний аналіз, аналіз відповідностей та інші методи для аналізу багатовимірних даних.
7. Непараметричний аналіз: Надає різноманітні непараметричні тести, включаючи вокальний тест, критерій серій, тест Фридмана та інші.
8. Часовий ряд і прогнозування: Допомагає в аналізі часових рядів, включаючи прогнозування, експоненціальне згладжування та аналіз тенденцій.
9. Статистичне управління процесами: Використовується для контролю якості та статистичного управління виробничими процесами.
10. Системний аналіз вимірів: Допомагає визначити оптимальні методи вимірювань та збору даних.
11. Дисперсійний аналіз: Використовується для визначення відмінностей між групами даних.

Minitab дозволяє користувачам проводити широкий спектр статистичних аналізів та досліджень, зокрема в областях контролю якості, наукових досліджень та ділового аналізу.

**MS Excel**, частина пакету Microsoft Office, є однією з найпоширеніших електронних таблиць та інструментів для обробки даних. Ця програма дійсно має свої переваги у багатьох областях, зокрема в бізнесі та обробці даних, але, як ви вказали, вона має свої обмеження в якості наукового інструмента. Ось деякі з плюсів та мінусів MS Excel у контексті наукового дослідження та статистичного аналізу:

Переваги MS Excel для статистичного аналізу:



1. Зручність в роботі: MS Excel має дуже інтуїтивний і легкий у використанні інтерфейс, що робить його доступним для широкого кола користувачів.

2. Ефективність в обробці та аналізі даних: Ви можете використовувати MS Excel для швидкої обробки даних, включаючи фільтрацію, сортування, обчислення середніх, медіани, стандартного відхилення та інших статистичних метрик.

3. Побудова базових діаграм: MS Excel має можливості для побудови простих графіків і діаграм, що може бути корисним для візуалізації даних.

4. Легка інтеграція з іншими програмами Microsoft Office: Ви можете вставляти таблиці та графіки Excel у документи Word та презентації PowerPoint для створення звітів і презентацій.

Мінуси MS Excel для статистичного аналізу:

1. Обмеження в статистичних функціях: Хоча Excel містить певні статистичні функції, він не має такого розширеного статистичного функціоналу, як спеціалізовані програми.

2. Якість наукових графіків: Графіки, які будуються в Excel, часто не відповідають стандартам якості для наукових публікацій.

3. Не підходить для складних аналітичних завдань: Для складних статистичних аналізів і моделювання, таких як аналіз часових рядів, байєсівський аналіз або машинне навчання, потрібні спеціалізовані інструменти.

4. Питання надійності та авторитетності: У науковому дослідженні часто вимагається, щоб аналіз був проведений відповідно до стандартів та був повністю документований. MS Excel не завжди відповідає цим вимогам.

У багатьох випадках великі організації та дослідницькі інститути використовують спеціалізовані статистичні програми, такі як SPSS, R або SAS, для більш складних наукових аналізів та досліджень. Проте Excel може бути корисним інструментом для початкового аналізу даних та невеликих завдань.

**NCSS** (Number Cruncher Statistical System) - це статистичний пакет, спрямований на використання користувачами без глибоких знань в області статистики, але які потребують статистичний аналіз та побудову графіків для своєї роботи або досліджень. Ця програма надає кілька переваг:

1. Спрощений інтерфейс: NCSS розроблений з урахуванням користувачів, які не мають глибоких знань в статистиці. Інтерфейс програми призначений для спрощення процесу введення даних, вибору відповідних статистичних методів та побудови графіків.

2. Багатовіконний інтерфейс: Багатовіконний підхід дозволяє вам працювати з кількома вікнами одночасно, що полегшує перегляд та порівняння різних аспектів даних та результатів аналізу.

3. Підказки та документація: NCSS надає підказки під час користування програмою, що полегшує розуміння статистичних методів та їх застосування. Користувачі також можуть скористатися документацією для докладнішого розуміння аналізу.

4. Побудова графіків: Програма дозволяє побудовувати графіки для візуалізації даних та результатів статистичного аналізу. Графіки можуть бути важливими для розуміння шаблонів та взаємозв'язків в даних.

5. Імпорт та експорт даних: NCSS підтримує імпорт та експорт даних у різних форматах, що дає можливість легко обробляти дані, отримані з інших джерел, та обмінюватися ними.

Хоча NCSS розрахований на користувачів, які не є професіоналами в області статистики, він може бути корисним для проведення базового статистичного аналізу та візуалізації даних у наукових дослідженнях, особливо тих, які не вимагають використання складних статистичних методів.

**OriginPro**, розроблений компанією OriginLab Corp., є потужним пакетом для аналізу та візуалізації даних, а також для створення наукових графіків та проведення статистичних обчислень. Основні функції та можливості OriginPro включають:

1. Аналіз та обробка даних: OriginPro дозволяє виконувати широкий спектр обчислень, включаючи статистичний аналіз, обробку сигналів, чисельний аналіз даних, регресійний аналіз і багато інших статистичних методів. Це корисно для наукових досліджень, інженерії, медицини та багатьох інших галузей.

2. Побудова наукових графіків: OriginPro дозволяє створювати 2D і 3D графіки для візуалізації даних та результатів досліджень. Ви можете користуватися готовими шаблонами графіків або створювати власні графічні візуалізації.

3. Підтримка програмування: OriginPro надає можливість програмувати власні аналітичні методи та скрипти за допомогою мови програмування, включаючи C/C++.

4. Модулі та розширення: У склад OriginPro входять різноманітні модулі, такі як Peak Fitting Module, які розширюють функціонал програми для конкретних завдань аналізу.

5. Експорт та обмін даними: OriginPro дозволяє експортувати результати аналізу та графіку в різні формати, такі як PDF, EPS, WMF, TIFF, JPEG, GIF. Це полегшує поділ результатів досліджень з іншими дослідниками або включення графіків у документи.

6. Підтримка стандартів: OriginPro підтримує роботу з багатьма науковими та інженерними стандартами, що робить його корисним для професіоналів з різних галузей.

Загалом, OriginPro є корисним інструментом для наукових досліджень, інженерних досліджень та аналізу даних завдяки своїм функціональним можливостям та зручному інтерфейсу.

**RATS** (Regression Analysis of Time Series), розроблений Estima, Inc., є програмним забезпеченням для аналізу часових рядів, зокрема в області економетрики. Він надає користувачам засоби для статистичного аналізу та моделювання часових рядів і включає ряд корисних функцій:

1. Метод найменших квадратів (OLS): RATS дозволяє використовувати метод найменших квадратів для оцінки параметрів статистичних моделей, включаючи регресійні моделі.

2. Ймовірнісні та логістичні моделі: Програма підтримує оцінку ймовірнісних моделей та моделей регресії з логістичною функцією.

3. Зважений метод найменших квадратів (WLS): За допомогою вагових коефіцієнтів можна враховувати важливість окремих спостережень при оцінці параметрів моделі.

4. Інструментальні змінні: Для моделювання зв'язків із змінними, що можуть бути ендогенними, RATS підтримує інструментальні змінні.

5. Критерій Кочрена-Оркатта: Цей критерій використовується для визначення кількості лагів в моделі часового ряду.

6. Поліноміальні розподілені лаги: Дозволяють включити поліноміальні лаги у модель для врахування часових залежностей.

RATS є популярним інструментом для аналізу часових рядів у галузі економетрики та фінансів. Він надає користувачам можливість обчислювати та оцінювати статистичні моделі для прогнозування, аналізу даних і прийняття рішень на основі часових рядів.

**S-PLUS** є потужним інструментом для аналізу даних, розробленим AT & T Bell Labs і пізніше перейнятий компанією TIBCO Software Inc. Він має об'єктно-орієнтовану мову програмування та велику кількість статистичних і графічних функцій, що дозволяє користувачам виконувати різноманітні аналітичні завдання.

Основні функції і можливості S-PLUS включають:

1. Об'єктно-орієнтований підхід: S-PLUS використовує об'єктно-орієнтований підхід до програмування, що дозволяє легко створювати, обробляти і взаємодіяти з об'єктами даних.
2. Статистичний аналіз: Програма надає широкий спектр статистичних методів для аналізу даних, включаючи дескриптивний аналіз, регресію, класифікацію, кластеризацію та багато інших.
3. Графіка: S-PLUS дозволяє створювати високоякісні графіки для візуалізації даних, включаючи статичні та динамічні графіки.
4. Можливість програмування: Користувачі можуть розробляти власні функції та процедури для вирішення специфічних завдань.
5. Інтеграція з C і FORTRAN: S-PLUS підтримує інтерфейс з програмами, написаними на мовах програмування C і FORTRAN, що дозволяє розширювати його функціональність.
6. Користувацький інтерфейс: Має користувацький інтерфейс з системою меню для зручного взаємодії з програмою.

S-PLUS широко використовується в областях статистики, дослідження даних, економетрики, та інших наукових та технічних дисциплінах для проведення аналізу та моделювання даних.

**SAS** (Statistical Analysis System) є потужною інтегрованою системою для аналізу даних та статистики, розробленою компанією SAS Institute, Inc. Ця програмна платформа здатна вирішувати різноманітні завдання з обробки і аналізу даних. Основні функції та можливості SAS включають:

1. Аналіз даних: SAS надає широкий спектр статистичних методів для обробки та аналізу даних, включаючи дослідження даних, дисперсійний аналіз, тестування гіпотез, регресійний аналіз, кластерний аналіз і багато інших.

2. Побудова звітів: SAS дозволяє створювати звіти та презентації на основі аналізу даних з використанням графіків, таблиць і текстових даних.

3. Підключення до баз даних: SAS має можливість взаємодіяти з різними системами управління базами даних, такими як Oracle і Ingres, для витягування даних та подальшого аналізу.

4. Аналіз часових рядів: SAS включає інструменти для аналізу часових рядів і прогнозування, що допомагають виявляти тенденції та прогнозувати майбутні значення.

5. Графіка: Програма дозволяє створювати високоякісні кольорові графіки для візуалізації даних та результатів аналізу.

6. Матричне програмування: SAS підтримує роботу з матрицями і векторами, що робить його відмінним інструментом для обчислень.

7. Експертна підтримка: SAS надає фахівцям можливість розробки власних алгоритмів та моделей для аналізу даних.

SAS використовується в різних галузях, включаючи бізнес-аналітику, епідеміологію, фінанси, медичинські дослідження, інженерію та багато інших. Це популярний інструмент серед професіоналів, які займаються аналітикою та дослідженнями даних.

**IBM SPSS Statistics**, коли-то відомий як Statistical Package for the Social Sciences (SPSS), є одним із найпопулярніших і потужних статистичних пакетів обробки даних. Основні характеристики SPSS включають:

1. Аналіз даних: SPSS надає широкий спектр статистичних методів для проведення різноманітних аналізів, включаючи дослідження, дисперсійний аналіз, кореляцію, регресію, кластерний аналіз і багато інших.

2. Побудова звітів: Програма дозволяє користувачам створювати звіти на основі результатів аналізу даних, використовуючи графіки, таблиці і текст.

3. Маркетингові дослідження: SPSS має модулі та функції, спеціально призначені для проведення маркетингових досліджень, включаючи аналіз маркетингових даних та прогнозування.

4. Інтеграція з іншими даними: SPSS може взаємодіяти з різними форматами даних, включаючи імпорт та експорт даних із зовнішніх джерел.

5. Візуалізація даних: Програма дозволяє створювати графіки та діаграми для візуалізації даних.

6. Зручний інтерфейс: SPSS має інтуїтивний та дружній інтерфейс, що полегшує роботу з програмою.

7. Аналіз великих об'ємів даних: SPSS може обробляти великі об'єми даних і виконувати аналіз, що полегшує обробку і розуміння великих наборів інформації.

8. Редагування даних: SPSS включає редактор даних для підготовки і очищення даних перед аналізом.

SPSS використовується в різних галузях, включаючи соціальні науки, бізнес-аналітику, медичнські дослідження і багато інших. Він допомагає професіоналам та дослідникам проводити аналіз даних та виводити інсайти із них. Хоча SPSS є комерційним програмним забезпеченням і вимагає ліцензійну плату, воно надає широкий набір інструментів для роботи з даними та аналізу, що робить його дуже популярним серед користувачів у багатьох сферах.

**STADIA** - це статистичний пакет, який має свої особливості та можливості в області аналізу даних. Основні характеристики та функції STADIA включають:

1. Ділова графіка: STADIA надає можливість побудови різних видів графіків та діаграм, таких як функції, залежності, розподіли, діаграми розсіяння, гістограми, стовпчикові, вежі, кругові діаграми і багато інших.

2. Статистичний аналіз: Програма містить різноманітні статистичні функції та методи, включаючи параметричні тести (наприклад, критерії Стюдента та

Фішера), непараметричні тести (наприклад,  $\chi^2$ -квадрат тест), аналіз часових рядів (згладжування, автокореляція, спектральний аналіз), дисперсійний аналіз (одно-, дво-, багатofакторний) та інші.

3. Категоріальний аналіз: STADIA дозволяє виконувати категоріальний аналіз, включаючи кростабуляцію та  $\chi^2$ -квадрат оцінку.

4. Багатовимірні методи: Програма підтримує багатовимірні методи аналізу, такі як дискримінантний аналіз, кластерний аналіз та факторний аналіз.

5. Редактор даних: STADIA включає редактор даних для підготовки та очищення даних перед аналізом.

6. Імпорт і експорт даних: Програма підтримує імпорт та експорт даних в різноманітних форматах, включаючи ASCII і DBF.

7. Контекстна екранна допомога: STADIA має розвинуту систему контекстної екранної допомоги для користувачів.

Хоча STADIA може бути потужним інструментом для статистичного аналізу даних, варто зазначити, що програма має свої обмеження, такі як застарілий зовнішній інтерфейс та колірну гамму, що може виглядати архаїчно.

**Stata** (Stata Corp.) - це потужний пакет для обробки статистичних та графічних даних, який призначений для фахівців і дослідників, займаються науковими дослідженнями в різних галузях знань, включаючи суспільні науки (економіка, політологія), медицину (біостатистика, епідеміологія) та інші. Важливі можливості і характеристики Stata включають:

1. Різноманітні моделі: Stata дозволяє користувачам застосовувати різноманітні моделі для статистичного аналізу даних, включаючи динамічні факторні моделі, лінійні та нелінійні моделі, узагальнені лінійні моделі, багаторівневі змішані моделі, моделі вибуття, узагальнені оцінки рівнянь (GEE) та інші.



2. Аналіз повторних вибірок: Stata підтримує аналіз повторних вибірок і статистичний експеримент, що дозволяє вивчати зміни в часі та вплив впроваджених змін на результати дослідження.

3. Перевірка гіпотез і робота з моделями: Користувачі можуть виконувати статистичні тести для перевірки гіпотез і роботи з оціненими моделями.

4. Максимізація функцій правдоподібності: Програма дозволяє максимізувати функції правдоподібності, які задаються користувачем для вирішення різних завдань.

5. Матричні команди: Stata має матричні команди, що полегшують обробку та аналіз даних, особливо в масштабних дослідженнях.

Це програмне середовище дозволяє дослідникам ефективно проводити аналіз даних, створювати статистичні моделі та виконувати різні види тестів. Stata також дозволяє виконувати продуктивну і розширену роботу з даними, включаючи імпорт, експорт та очищення даних.

**Statistica**, розроблена компанією StatSoft, Inc., є добре збалансованим статистичним пакетом, який відзначається комбінацією потужних функцій та зручного інтерфейсу. Основні характеристики і переваги Statistica включають:

1. Широкий функціонал: Statistica має понад 250 статистичних функцій та алгоритмів, які дозволяють проводити різноманітний статистичний аналіз даних, включаючи аналіз часових рядів, дисперсійний аналіз, регресію, непараметричну статистику, кластерний аналіз, факторний аналіз і багато іншого.

2. Зручний інтерфейс: Statistica має інтуїтивно зрозумілий і легкий у використанні інтерфейс. Це робить його доступним для широкого кола користувачів, включаючи тих, хто не є професійними статистиками.

3. Підтримка російської мови: Наявність російської версії програми та російськомовного сайту підтримки робить Statistica зручною для користувачів, які володіють російською мовою.

4. Можливості програмування: Statistica має вбудовані мови програмування, такі як SCL (Statistica Command Language) і STATISTICA BASIC, що дозволяють створювати власні скрипти та автоматизувати аналітичні завдання.

5. Розширення функціональності: Statistica дозволяє користувачам розширювати бібліотеку функцій шляхом написання власних модулів, що дозволяє вирішувати більше завдань.

6. Сумісність з іншими програмами: Statistica може обмінюватися даними з іншими програмами, включаючи Microsoft Office Word та інші, що полегшує обробку і представлення результатів.

Загалом, Statistica є потужним і універсальним інструментом для проведення аналізу даних та статистичних досліджень у різних галузях знань.

**StatGraphics**, розроблений компанією Manugistics, Inc., є універсальним статистичним пакетом, який надає можливості для діалогового аналізу даних та високоякісної візуалізації результатів. Основні характеристики та переваги StatGraphics включають:

1. Регресійний аналіз: StatGraphics дозволяє проводити різні види регресійного аналізу, включаючи просту регресію, багатофакторну лінійну та нелінійну регресію, а також авторегресійні моделі. Це корисно для прогнозування та моделювання залежностей між змінними.

2. Аналіз часових рядів: StatGraphics підтримує аналіз часових рядів, включаючи авторегресійні моделі та інші методи аналізу, які корисні для роботи з даними, що змінюються в часі.

3. Візуалізація даних: StatGraphics відзначається високоякісною двовимірною та тривимірною графікою. Ви можете створювати інтерактивні графіки та графічні представлення результатів аналізу, що допомагає вам краще розуміти ваші дані та представляти їх іншим.

4. Взаємодія з іншими програмами: StatGraphics підтримує взаємодію з іншими програмами, такими як електронні таблиці та бази даних. Це полегшує імпорт і експорт даних та спільну роботу з іншими інструментами.

5. Обрання графічного відображення результатів: Після проведення статистичного аналізу даних ви можете обирати графічні відображення результатів, які найкраще підходять для вашої конкретної задачі. Це сприяє зрозумілій інтерпретації результатів.

Загалом, StatGraphics є корисним інструментом для аналізу даних та візуалізації результатів, особливо в тих випадках, коли важлива висока якість графічного представлення даних.

**SYSTAT**, розроблений компанією Systat Software, Inc., є програмою для статистичної обробки та аналізу даних з багатьма корисними функціями. Основні можливості SYSTAT включають:

1. Ділова графіка: SYSTAT надає інструменти для створення якісних графічних представлень даних, включаючи графіки, таблиці, гістограми та інші графічні засоби.

2. Параметричні та непараметричні тести: Програма підтримує проведення різних статистичних тестів, включаючи критерії Стьюдента, Фішера, а також непараметричні тести, такі як тест Колмогорова-Смірнова, хі-квадрат та інші.

3. Категоріальний аналіз: SYSTAT дозволяє проводити аналіз категоріальних даних, включаючи кростабуляцію та оцінку коефіцієнтів узгодження між категоріями.

4. Дисперсійний та регресійний аналіз: Ви можете використовувати програму для проведення дисперсійного аналізу, включаючи одно-, дво-, та багатofакторний аналіз, а також регресійний аналіз, включаючи просту та множинну регресію.

5. Аналіз часових рядів: SYSTAT підтримує аналіз часових рядів, включаючи методи згладжування, фільтрації, автокореляції та ARIMA-моделі.

6. Багатовимірні методи: Програма дозволяє проводити аналіз багатовимірних даних, включаючи кореляцію, кластерний аналіз та факторний аналіз.

7. Глибокий аналіз цифрових зображень: SYSTAT також має можливості для аналізу та обробки цифрових зображень.

Ця програма надає широкий спектр статистичних і аналітичних інструментів та є корисною для проведення різних видів статистичного аналізу і обробки даних.

**Unistat Statistical Package**, розроблений Unistat Statistics Software, є програмою для аналізу статистичних даних та результатів наукових експериментів. Основні особливості цього статистичного пакету включають:

1. Повний набір функцій: Unistat надає розширений набір статистичних і аналітичних функцій, які дозволяють вам проводити різні види аналізу даних, включаючи діловий аналіз, параметричні та непараметричні тести, регресійний аналіз, кластерний аналіз та інші.

2. Інтеграція з Excel: Unistat інтегрований з Microsoft Excel, що дозволяє вам легко використовувати статистичні можливості в Excel. Ви можете використовувати Unistat як плагін для Excel для аналізу даних безпосередньо в Excel.

3. Графічні можливості: Програма надає можливості створення графічних представлень даних, включаючи графіки, гістограми, діаграми розсіювання та інші типи графіків для ілюстрації результатів аналізу.

4. Аналіз часових рядів: Unistat підтримує аналіз часових рядів, включаючи методи згладжування, прогнозування та аналіз трендів.

5. Легкий використання: Програма має користувацький інтерфейс, що робить її доступною для користувачів з різним рівнем досвіду в аналізі даних та статистиці.

Unistat дозволяє проводити статистичний аналіз даних у зручному середовищі Excel та надає різноманітні інструменти для аналізу та візуалізації даних. Інтеграція з Excel робить його популярним серед користувачів, які звикли працювати з цією електронною таблицею.

Система **ЭВРИСТА** (розроблена ООО «Центр Статистических Исследований») є статистичним програмним забезпеченням з різноманітними функціями аналізу даних. Основні можливості цієї системи включають:

1. Ділова графіка: Система дозволяє створювати графіки та візуалізації для ілюстрації результатів аналізу даних.

2. Параметричні тести: Включають в себе описову статистику та гістограми для аналізу розподілу даних.

3. Непараметричні тести: Серед них хі-квадрат тест, тест знаків, тест Колмогорова-Смірнова, які використовуються для аналізу незалежних вибірок та розподілу даних.

4. Регресійний аналіз: Включає просту регресію та множинну регресію для дослідження залежностей між змінними.

5. Аналіз часових рядів: Система має багато функцій для аналізу часових рядів, включаючи згладжування, фільтрацію, автокореляцію, кроскореляцію, спектральний аналіз, крос-спектральний аналіз та ARIMA-моделі.

6. Багатовимірні методи: Включає аналіз факторів та факторний аналіз для розкриття залежностей між багатьма змінними.

Згадується, що перша версія системи ЭВРИСТА була реалізована мовою ФОРТРАН для ЕОМ БЭСМ-6. Дане програмне забезпечення, ймовірно, було створене для використання в сфері статистичних аналізів і досліджень, і може бути корисним для аналізу даних і моделювання залежностей між різними факторами у відповідному контексті.

## **Висновки до 2 розділу**

В даному розділі було розглянуто сучасне програмне забезпечення для використання багатовимірного статистичного аналізу даних. Ми проаналізували ринок програмного забезпечення та сучасні тенденції розвитку. Також, ми розглянули найкращих представників програмних засобів для математичної і статистичної обробки даних. Дослідили та порівняли характеристики програмних пакетів, розглянули переваги та їх недоліки.

## РОЗДІЛ 3. АЛГОРИТМ І КОМП'ЮТЕРНА РЕАЛІЗАЦІЯ ЗАДАЧІ ПЕРЕВІРКИ ОДНОРІДНОСТІ БАГАТОВИМІРНИХ СТАТИСТИЧНИХ ДАНИХ

### 3.1 РОЗРАХУНКОВІ СПІВВІДНОШЕННЯ ТА ПРОГРАМНЕ ЗАБЕЗПЕЧЕННЯ

З матеріалу, що був розміщений у попередніх розділах, випливає, що важливою частиною аналізу будь-яких даних (числового, векторного, якісного характеру) є перевірка гіпотези про їхню однорідність. Однорідність даних свідчить про їх спільне походження або ідентичність (у певному сенсі) джерел, що генерують зазначені дані. Зазначене питання може мати багато аспектів: наявність аномальних даних, значуща різниця середніх значень різних вибірок, відмінність коваріацій між компонентами векторних статистик для різних джерел, більш загально, підпорядкованість різних груп даних різним ймовірнісним розподілам та інше. В даний час розроблено набір методів, що дозволяє виконувати належні дії по перевірці наявності зазначених ситуацій. Наприклад, для виявлення аномальних спостережень і відмінності коваріацій використовуються статистики Махаланобіса і Бартлета [27,28]. Вони будуються у спосіб, що наводиться нижче.

#### Деякі способи перевірки однорідності даних

##### 1) Перевірка присутності аномальних даних у векторній вибірці

Зазначена процедура заснована на обчисленні так званої *вибіркової відстані Махаланобіса*  $D^2$ , явний вираз якої дається формулою (див., напр., [27]):

$$D^2 = (x - \bar{x})' S^{-1} (x - \bar{x}),$$

в якій

$$\bar{x} = \frac{1}{k} \sum_{i=1}^k x_i, \quad S = \frac{1}{k-1} \sum_{i=1}^k (x_i - \bar{x})(x_i - \bar{x})'.$$



$x_i, i = 1, \dots, k$  — векторні вибіркові дані розмірності  $p$ , вже перевірені на наявність аномалій,  $x$  — вектор, що перевіряється. У випадку відсутності аномалій статистика

$$F = \frac{(k-p)k}{(k^2-1)p} D^2$$

має розподіл Фішера  $F_{p,k-p}$ . При заданому  $\alpha$  гіпотеза про аномальність  $x$  приймається при перевищенні статистикою  $F$  квантиля рівня  $1 - \alpha$  зазначеного розподілу. Підкреслимо, що програма обчислює такі квантили, і звертатися до додаткового забезпечення чи до статистичних таблиць немає потреби.

## 2) Перевірка гіпотези про рівність кількох коваріаційних матриць

Нехай одержано  $k$  незалежних вибірок об'ємів  $n_1, n_2, \dots, n_k$  з  $p$ -вимірних нормальних сукупностей з середніми  $\mu_1, \mu_2, \dots, \mu_k$  і коваріаційними матрицями  $\Sigma_1, \Sigma_2, \dots, \Sigma_k$  відповідно. Для багатьох статистичних процедур, зокрема для виконання дій дисперсійного аналізу, важливо знати, чи виконується гіпотеза  $H_0: \Sigma_1 = \Sigma_2 = \dots = S$ . Критерій Бартлета та його узагальнення, що розглядається у даній роботі, заснований на статистиці перевірки  $M$ , вираз якої є наступним:

$$M = (n - k) \ln |S_U| - \sum_{i=1}^k (n_i - 1) \ln |S_{U_i}|, \quad \sum_{i=1}^k n_i = n,$$

де  $S_{U_i}$  — незміщена оцінка коваріаційної матриці  $S_i$ , що побудована за  $i$ -ю вибіркою,  $S_U = \frac{1}{n-k} \cdot \sum_{i=1}^k (n_i - 1) \cdot |S_{U_i}|$  та  $|S_{U_i}|$  — детермінанти зазначених матриць.

Гіпотеза  $H_0$  відкидається при великих значеннях статистики  $M$ . В силу того, що нульовий розподіл статистики  $M$  при справедливості гіпотези  $H_0$  є досить складним, він апроксимується  $F$ -розподілом, оснований на співвідношенні

$$Mb^{-1} \cong F_{v_1, v_2},$$

$$\text{де } A_1 = \frac{2p^2+3p-1}{6(k-1)(p+1)} \left( \sum_{i=1}^k \frac{1}{n_i-1} - \frac{1}{n-k} \right), A_2 = \frac{(p-1)(p+2)}{6(k-1)} \left( \sum_{i=1}^k \frac{1}{(n_i-1)^2} - \frac{1}{(n-k)^2} \right)$$

$$v_1 = \frac{(k-1)p(p+1)}{2}, v_2 = \frac{v_1+2}{A_2-A_1^2}, b = \frac{v_1}{1-A_1-v_1/v_2}.$$

Зауважимо, що велику кількість способів виявлення неоднорідності або аномальності вибірових даних (необов'язково числових) дають процедури так званого розвідувального аналізу [29].

Комп'ютерне програмне забезпечення, що розроблено в даній кваліфікаційній роботі, відноситься до перевірки даних (довільного, зокрема економічного характеру) на однорідність коваріаційних зв'язків. Розроблена програма автоматизує дії по застосуванню вищезгаданого критерію Бартлета. Зауважимо, що можливість інтерпретувати його результати об'єктивним чином залежить від підпорядкованості даних векторному (багатовимірному) розподілу імовірностей. Наведемо основні означення з даного приводу.

### **Векторний) нормальний розподіл імовірностей та деякі його властивості**

У даному розділі подаються визначення та інформація про властивості нормального розподілу ймовірностей  $N_p$  в скінченновимірному евклідовому просторі  $R_p$ . Для ознайомлення з доведенням наведених фактів, а також більш фундаментальним викладенням матеріалу, можна звернутися до джерел, таких як [30,31]. Визначимо  $0$  як нульовий вектор, а  $I$  як одиничну матрицю:

$$\mathbf{0} = \begin{pmatrix} 0 \\ 0 \\ \dots \\ 0 \\ 0 \end{pmatrix}, \quad \mathbf{I} = \begin{pmatrix} 1 & 0 & \dots & 0 & 0 \\ 0 & 1 & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}.$$

Згідно з вищезазначеними позначеннями, при необхідності уточнення їх розмірностей, використовуються відповідні індекси, наприклад  $\mathbf{0}_p, \mathbf{I}_{p \times p}$ .

**3.1 Визначення.** Позначимо  $N_p(\mathbf{0}, \mathbf{I})$  закон розподілу  $p$  - вимірною випадкового вектора, складові якого є взаємно незалежними нормально розподілені випадкові величини з параметрами  $(0,1)$  кожна. Такий вектор отримує назву стандартного  $p$ -вимірною нормального вектора.

Отже, у позначенні цього закону нижній індекс  $p$  вказує на вимірюваність відповідного випадкового вектора. Якщо ця вимірюваність чітко визначається контекстом, то цей індекс може бути відсутнім, тобто у цьому випадку використовується позначення  $N(\mathbf{0}, \mathbf{I})$ . Зокрема, при  $p = 1$  розподіл  $N(\mathbf{0}, \mathbf{I}) = N_1(\mathbf{0}, \mathbf{I})$  є скалярним розподілом  $N(0,1)$ , тобто одновимірним нормальним розподілом з параметрами  $0,1$ .

З визначення випливає, що вектор  $\eta$ , розподілений за законом  $N_p(\mathbf{0}, \mathbf{I})$ , має нульовий  $p$  - вимірний вектор математичного сподівання і коваріаційну матрицю, котра є просто одиничною матрицею розмірності  $p \times p$ :

$$M \xi = \mathbf{0} = \mathbf{0}_p, \quad D\eta = \mathbf{I} = \mathbf{I}_{p \times p}. \quad (3.1)$$

**3.2 Визначення.** Нехай вектор  $\eta$  має розподіл  $N_p(\mathbf{0}, \mathbf{I})$ , а вектор  $\xi$  визначається з рівності

$$\xi = A\eta + \mu, \quad (3.2)$$

в якій  $A$  — матриця розмірності  $q \times p$ ,  $\mu$  — вектор розмірності  $q$ . Розподіл вектора  $\xi$  називається загальним  $q$ -вимірним нормальним розподілом.

Для  $\xi = (\xi_1, \xi_2, \dots, \xi_q)'$  синонімами висловлювання «вектор  $\xi$  має нормальний розподіл» є «вектор  $\xi$  розподілений за нормальним законом» або «випадкові величини  $\xi_1, \xi_2, \dots, \xi_q$  мають сумісний нормальний розподіл». Можуть бути і деякі інші (ідентичні за значенням) варіанти.

Покладемо  $\Sigma = AA'$  і далі будемо позначати тільки що означений розподіл  $N_q(\mu, \Sigma)$ . Також будемо писати просто  $N(\mu, \Sigma)$  (без нижнього індексу), якщо розмірність не викликає питань або не є суттєвою.

З рівності (3.1) і відомих властивостей середніх значень та коваріаційно-дисперсійних матриць випадкових векторів випливає, що для вектора  $\xi$  класу  $N(\mu, \Sigma)$  мають місце рівності:

$$M\xi = \mu, D\xi = \Sigma. \quad (3.3)$$

Зауважимо, що в позначенні  $N(\mu, \Sigma)$  існує певна відмінність від стандартного позначення  $N(\mu, \sigma)$  для нормального розподілу скалярних випадкових величин. Ця відмінність полягає в тому, що в останньому випадку  $\sigma$  позначає середнє квадратичне відхилення  $\xi$ , тобто  $(D\xi)^{1/2}$ , тоді як у виразі  $N(\mu, \Sigma)$  при  $q = 1$  символ  $\Sigma$  позначає дисперсію, тобто  $D\xi$ .

Сучасне програмне забезпечення, зокрема, комплекси статистичних програм, багато уваги приділяє питанням оцінювання параметрів і перевірки статистичних гіпотез у випадку вибірок даних числового характеру. Питання виконання аналогічних дій для векторних (багатовимірних) даних представлені у названих джерелах значно менш повно, і досить часто без належних роз'яснень з приводу використання можливостей, що мають місце. Враховуючи сказане, автори статті розробили комп'ютерну програму, що розв'язує наступні задачі.

### 3.3 Про збереження нормальності розподілів випадкових векторів при афінних перетвореннях.

Відзначимо важливу властивість нормально розподілених випадкових векторів.

Якщо  $\xi \in N_q(\mu, \Sigma)$ , а  $r$ -вимірний вектор  $\zeta$  є афінним перетворенням  $\xi$ , тобто

$$\zeta = B\xi + v, \quad (3.4)$$

У випадку, коли матриця  $B = B_{r \times q}$  — є не випадковою матрицею розмірності  $(r \times q)$ - а,  $v$  представляє собою  $r$ -вимірний не випадковий вектор, то отриманий вектор  $\zeta$  також має нормальний розподіл, конкретно,  $\zeta \in N_r(\lambda, K)$ , де  $\lambda = B\mu + v$ ,  $K = B \Sigma B'$ .

**3.4 Наслідок.** 1) Лінійне перетворення вектора, який має нормальний розподіл, само є вектором з нормальним розподілом. 2) Точніше, якщо  $\zeta = B\xi$ , де  $\xi$  належить  $N(\mu, \Sigma)$ , при умові, що  $B$  — певна матриця (за розмірністю, яка відповідає  $\xi$ ), то  $\zeta$  належить  $N(B\mu, B \Sigma B')$ .

**3.5 Наслідок.** 1) Сумісний розподіл будь-якої підмножини компонент нормального випадкового вектора також є нормальним випадковим вектором.

2) Детальніше, якщо  $q$ -вимірний вектор  $\xi$  має розподіл  $N(\mu, \Sigma)$ , і вектор  $\xi^j$  формується видаленням його компонентів з номерами  $j_1, j_2, \dots, j_k$ ,  $1 \leq k < q$ , то  $(q - k)$ -вимірний вектор  $\xi^j$  має розподіл  $N(\mu^j, \Sigma^j)$ , де вектор  $\mu^j$  формується з вектора  $\mu$  аналогічно до того  $\xi^j$  з  $\xi$ , а матриця  $\Sigma^j$  утворюється з матриці  $\Sigma$  видаленням рядків і стовпців з тими самими номерами  $j_1, j_2, \dots, j_k$ .

**3.6 Наслідок.** Нехай  $\xi = (\xi_1, \xi_2, \dots, \xi_q)'$  має розподіл  $N(\mu, \Sigma)$ ,  $c = (c_1, c_2, \dots, c_q)'$  це дійсний вектор (невипадковий). Тоді лінійна комбінація  $c_1\xi_1 + c_2\xi_2 + \dots + c_q\xi_q$  має

(одновимірний) нормальний розподіл з середнім значенням  $c'\mu$  і дисперсією  $c'\Sigma c$ .

**3.7 Наслідок.** Якщо величини  $\xi_1, \xi_2, \dots, \xi_q$  мають сумісний нормальний розподіл  $N(\mu, \Sigma)$  ( $\mu = (\mu_1, \dots, \mu_q)'$ ,  $\Sigma = [\sigma_{ij}]_{q \times q}$ ), то їхня сума  $\zeta$  також є випадковою величиною з нормальним розподілом. Математичне сподівання та дисперсія  $\zeta$  задаються за допомогою рівностей:

$$M\zeta = \sum_{j=1}^q \mu_j, \quad D\zeta = \sum_{j=1}^q \sigma_{jj}^2 + 2 \sum_{i,j,i < j} \sigma_{ij}. \quad (3.5)$$

Дійсно, слід застосувати наслідок 1.6 при  $c = c_q = (1, 1, \dots, 1)'$ .

Відомо (див. наприклад, [3,6,7]), що однією з найважливіших ймовірнісних характеристик випадкових величин і векторів є їхні характеристичні функції  $\varphi_{\xi}(t)$ :

Відомо: *характеристична функція  $\varphi_{\xi}(t)$  випадкового вектора  $\xi \in N(\mu, \Sigma)$  відповідає рівності*

$$\varphi_{\xi}(t) = e^{it\mu - \frac{t'\Sigma t}{2}} \quad (3.6)$$

У даному контексті вектор  $\mu$  — представляє собою вектор середнього значення для випадкового вектора  $\xi$  (математичне сподівання), а  $\Sigma$  — коваріаційна матриця  $\xi$ . Важливо відзначити, що вектор  $\mu$  у рівності (3.6) може мати будь-яке значення, тоді як матриця  $\Sigma$ , як і будь-яка коваріаційна матриця, повинна бути невід'ємно визначеною.

У протилежність цьому, для будь-якого значення вектора середнього  $\mu$  та довільної невід'ємно визначеної матриці коваріації  $\Sigma = \Sigma_{q \times q}$ , **вираз (3.4)** є

представленням характеристичної функції для конкретного  $q$ -вимірного нормального розподілу ймовірностей.

Наступна властивість є характеристичною для випадкових векторів, що розподілені нормально. Оскільки ця властивість є основою для запропонованого у роботі методу перевірки нормальності векторного розподілу, то в теоремі 3.8 наводиться і доведення цього результату.

Теорема 3.8 стверджує, що для того, щоб  $q$ -вимірний випадковий вектор  $\xi$  мав нормальний розподіл, необхідно і достатньо, щоб для будь-якого дійсного  $q$ -вимірного невикладкового вектора  $c$  випадкова величина  $c'\xi$  була нормально розподіленою випадковою величиною (за означенням, одновимірною).

Необхідність цієї теореми є частковим випадком наслідку 3.6. Доведемо достатність. З нормальності (одновимірної) величини  $c'\xi$  випливає, що її характеристична функція  $\varphi_{c'\xi}(t)$  задається рівністю

$$\varphi_{c'\xi}(t) = \exp \{ itM(c'\xi) - t^2 D(c'\xi)/2 \}, t \in (-\infty, +\infty).$$

З різноманіття векторів  $c$ , зокрема, видно, що кожна компонента вектора  $\xi$  як нормально розподілена випадкова величина, має обмежені перший і другий моменти. Це дозволяє зробити висновок, що кожен елемент коваріаційної матриці вектора  $\xi$  є визначеним і скінченим. Позначимо  $\mu$  вектор математичного сподівання  $\xi$ , а  $\Sigma$  — його кореляційну матрицю. Тоді з останньої рівності і рівності (1.2.6) отримуємо, що

$$\varphi_{c'\xi}(t) = \exp \{ it(c'\mu) - t^2(c'\Sigma c)/2 \}.$$

З іншого боку, оскільки  $\varphi_{c'\xi}(t) = M e^{it(c'\xi)}$ , то  $\varphi_{c'\xi}(1) = M e^{i(c'\xi)}$ . Але останній вираз дорівнює характеристичній функції  $\varphi_{\xi}$  вектора  $\xi$  в точці  $c$ . Тому, покладаючи  $t = 1$ ,  $\forall c \in R^q$  маємо рівність

$$\varphi_{\xi}(\mathbf{c}) = \exp \{ i(\mathbf{c}'\boldsymbol{\mu}) - (\mathbf{c}'\boldsymbol{\Sigma}\mathbf{c})/2 \},$$

права частина якої як функція від  $\mathbf{c}$  представляє собою характеристичну функцію нормального розподілу.

Розглянемо тепер питання існування щільності розподілу вектора  $\xi \in N_q(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ . Якщо ранг дисперсійної матриці  $\boldsymbol{\Sigma}$  менший за  $q$ , то можна показати, що розподіл  $\xi$  концентрований у власній підмножині простору  $R^q$ , і щільність розподілу у цьому випадку не існує. Припустимо, що ранг  $\boldsymbol{\Sigma}$  дорівнює  $q$ . У цьому випадку щільність  $f_{\xi}$  існує. Тепер ми переконаємося у цьому і знайдемо її вираз. Позначимо знову  $\boldsymbol{\eta}$  — стандартний нормальний  $q$ -вимірний вектор, пов'язаний з  $\xi$  рівністю (3.2), в якому матриця  $\mathbf{A} = \mathbf{A}_{q \times q}$  має ранг  $q$  і задовольняє рівності (3.7). Щільність розподілу  $f_{\boldsymbol{\eta}}$  знаходиться як добуток одновимірних щільностей компонентів вектора  $\boldsymbol{\eta}$ , тобто

$$f_{\boldsymbol{\eta}}(\mathbf{x}) = \prod_{j=1}^q \left( \frac{1}{\sqrt{2\pi}} e^{-\frac{x_j^2}{2}} \right) = (2\pi)^{-\frac{q}{2}} e^{-\frac{x_1^2 + \dots + x_q^2}{2}} = (2\pi)^{-\frac{q}{2}} e^{-\frac{1}{2}\mathbf{x}'\mathbf{x}} \quad (3.7).$$

Для одержання щільності  $f_{\xi}(\mathbf{x})$  треба в останньому виразі зробити такі перетворення: помножити його на  $|\det \mathbf{A}^{-1}|$  і замінити  $\mathbf{x}$  на  $\mathbf{A}^{-1}(\mathbf{x} - \boldsymbol{\mu})$ . При цьому з рівності (3.7) бачимо, що  $|\det \mathbf{A}^{-1}| = (\det \boldsymbol{\Sigma})^{-1/2}$ , а показник експоненти після зазначеної заміни з точністю до множника  $(-1/2)$  буде дорівнювати

$$\begin{aligned} (\mathbf{A}^{-1}(\mathbf{x} - \boldsymbol{\mu}))'(\mathbf{A}^{-1}(\mathbf{x} - \boldsymbol{\mu})) &= (\mathbf{x} - \boldsymbol{\mu})'(\mathbf{A}^{-1})' \mathbf{A}^{-1}(\mathbf{x} - \boldsymbol{\mu}) = (\mathbf{x} - \boldsymbol{\mu})' (\mathbf{A} \mathbf{A}')^{-1}(\mathbf{x} - \boldsymbol{\mu}) = \\ &= (\mathbf{x} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu}). \end{aligned}$$

Остаточно

$$f_{\xi}(\mathbf{x}) = (2\pi)^{-q/2} (\det \boldsymbol{\Sigma})^{-1/2} e^{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})'\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})}. \quad (3.8)$$



**3.9 Висновок.** Для існування щільності розподілу вектора  $\xi \in N(\mu, \Sigma)$  необхідно і достатньо, щоб коваріаційна матриця  $\Sigma$  була невідродженою. В цьому випадку щільність розподілу  $\xi$  задається рівністю (3.8).

**3.10 Двовимірний випадок.** Розглянемо спеціально двовимірний нормальний. Середнє значення  $\mu$  вектора  $\xi = (\xi_1, \xi_2)'$  дорівнює

$$M \begin{pmatrix} \xi_1 \\ \xi_2 \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \quad (3.9)$$

а матриця коваріацій  $\Sigma$  може бути записаною наступним чином:

$$\Sigma = M \begin{pmatrix} (\xi_1 - \mu_1)^2 & (\xi_1 - \mu_1)(\xi_2 - \mu_2) \\ (\xi_2 - \mu_2)(\xi_1 - \mu_1) & (\xi_2 - \mu_2)^2 \end{pmatrix} = \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{21} & \sigma_2^2 \end{pmatrix}, \quad (3.10)$$

де  $\sigma_1^2$  є дисперсією  $\xi_1$ ,  $\sigma_2^2$  — дисперсією  $\xi_2$ , а  $\sigma_{12} = \sigma_{21} = \text{cov}(\xi_1, \xi_2)$  (коваріація між  $\xi_1$  і  $\xi_2$ ). Позначимо  $\rho$  коефіцієнт кореляції між  $\xi_1$  і  $\xi_2$ :

$$\rho = \text{cov}(\xi_1, \xi_2) / (D\xi_1 D\xi_2)^{1/2} = \sigma_{12} / (\sigma_1 \sigma_2).$$

Тому коваріаційна матриця  $\Sigma$  вектора  $\xi$  може бути записаною у вигляді

$$\Sigma = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix}. \quad (3.11)$$

Неважно пересвідчитись, що матрицею, оберненою до матриці (3.11), є матриця

$$\Sigma^{-1} = \frac{1}{1-\rho^2} \begin{pmatrix} \frac{1}{\sigma_1^2} & -\frac{\rho}{\sigma_1\sigma_2} \\ -\frac{\rho}{\sigma_1\sigma_2} & \frac{1}{\sigma_2^2} \end{pmatrix}. \quad (3.12)$$

Щільність імовірності  $\xi$  (щільність сумісного розподілу  $\xi_1$  і  $\xi_2$ ), згідно із рівностями (3.8) і (3.12) дорівнює

$$f_{\xi}(x) = \frac{1}{2\pi\sigma_1\sigma_2} \exp\left\{-\frac{1}{2(1-\rho^2)} \left[ \frac{(x_1 - \mu_1)^2}{\sigma_1^2} - 2\rho \frac{(x_1 - \mu_1)(x_2 - \mu_2)}{\sigma_1\sigma_2} + \frac{(x_2 - \mu_2)^2}{\sigma_2^2} \right]\right\}. \quad (3.13)$$

При

$$\rho = 0 \quad (3.14)$$

щільність (3.13) розпадається на добуток двох щільностей одновимірних нормальних розподілів. Зокрема, за умовою (3.14) величини  $\xi_1$  і  $\xi_2$  є взаємно незалежними.

З даного виразу (3.13) можна визначити, що щільність нормального розподілу залишається постійною на кожному еліпсі

$$\frac{(x_1 - \mu_1)^2}{\sigma_1^2} - 2\rho \frac{(x_1 - \mu_1)(x_2 - \mu_2)}{\sigma_1\sigma_2} + \frac{(x_2 - \mu_2)^2}{\sigma_2^2} = c^2, \quad (3.15)$$

де  $c$  — константа. За іншими словами, еліпси (3.15) є лініями рівня щільності розподілу (3.13). Інколи їх називають еліпсами рівних ймовірностей.

У випадку загального  $q$ -вимірному розподілу поверхні рівня (еліпсоїди) нормальної щільності ймовірності (3.8) мають рівняння

$$(x - \mu)' \Sigma^{-1} (x - \mu) = c^2. \quad (3.16)$$

### Сумісні та умовні розподіли нормальних випадкових векторів.

**3.11 Означення.** Нехай  $\xi$  та  $\zeta$  —  $q$ -вимірний та  $r$ -вимірний випадкові вектори відповідно. Будемо говорити, що  $\xi$  та  $\zeta$  мають сумісний нормальний

розподіл (або що сумісний розподіл  $\xi$  та  $\zeta$  є нормальним), якщо  $(q + r)$ -вимірний вектор  $\gamma = \begin{pmatrix} \xi \\ \zeta \end{pmatrix} (= (\xi', \zeta'))'$  має нормальний розподіл.

Узагальнення означення сумісного нормального розподілу на випадок довільної скінченної кількості випадкових векторів, виконується очевидним чином.

**3.12 Зауваження.** З наслідку 1.5 випливає, що випадкові вектори, сумісний розподіл ймовірностей яких є нормальним, самі розподілені за нормальним законом.

З означення випливає, що характеристична функція  $\varphi_\gamma(t)$  вектора  $\gamma$  дорівнює

$e^{it'\mu - \frac{1}{2}t'\Sigma t}$ , де  $\mu = (\mu_1, \mu_2, \dots, \mu_{q+r})'$  —  $(q + r)$ -вимірний вектор,  $\Sigma$  — невід'ємно визначена матриця розміру  $(q + r) \times (q + r)$ . З того ж наслідку 1.5 вбачаємо, що середнє значення  $\mu_1$  вектора  $\xi$  дорівнює  $(\mu_1, \mu_2, \dots, \mu_q)'$ , а його кореляційна матриця, яку ми позначимо  $\Sigma_{11}$ , одержується з матриці  $\Sigma$  вилученням рядків і стовпців з  $(q + 1)$ -го по  $(q + r)$ -й включно. Аналогічно середнє значення  $\mu_2$  вектора  $\zeta$  дорівнює  $(\mu_q, \mu_{q+1}, \dots, \mu_{q+r})'$ , а кореляційна матриця  $\Sigma_{22}$  одержується вилученням рядків і стовпців  $\Sigma$  з першого по  $q$ -й включно. Маємо розбиття вектора  $\mu$  на складові «підвектори»:

$$\mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} \quad (3.17)$$

і, аналогічно, розбиття матриці  $\Sigma$  на складові «підматриці» (блоки):

$$\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}. \quad (3.18)$$

Тут  $\Sigma_{11}$  і  $\Sigma_{22}$  були визначені вище, а  $\Sigma_{12}$  і  $\Sigma_{21}$  — матриці, що доповнюють їх до  $\Sigma$  ( $\Sigma_{12}$  одержується з  $\Sigma$  видаленням перших  $q$  стовпців і останніх  $r$  рядків, а  $\Sigma_{21}$  — видаленням з  $\Sigma$  перших  $q$  рядків і останніх  $r$  стовпців).

Таким чином, вектор  $\xi$  має розподіл  $N_q(\mu_1, \Sigma_{11})$ , вектор  $\zeta$  — розподіл  $N_r(\mu_2, \Sigma_{22})$ , а матриці  $\Sigma_{12}$  і  $\Sigma_{21}$  характеризують кореляційні зв'язки між  $\xi$  і  $\zeta$  (з урахуванням співвідношень  $\sigma_{ij} = \sigma_{ji}$  легко бачити, що  $\Sigma_{21} = \Sigma_{12}'$ ).

**3.13 Теорема.** Для того, щоб вектори  $\xi$  і  $\zeta$ , що мають сумісний нормальний розподіл, були незалежними, необхідно та достатньо, щоб матриця  $\Sigma_{12}$  з рівності (3.18) була нульовою:

$$\Sigma_{12} = \mathbf{O}_{q \times r} \quad (3.19)$$

(останню рівність, звичайно, можна записати і як  $\Sigma_{21} = \mathbf{O}_{r \times q}$ ).

Оскільки коваріація незалежних випадкових величин дорівнює 0, необхідність є досить очевидною. Для того, щоб довести достатності помітимо, що за виконанням умови (3.19) характеристична функція  $\varphi_\gamma$  вектора  $\gamma = \begin{pmatrix} \xi \\ \zeta \end{pmatrix}$  розпадається у добуток характеристичних функцій  $\varphi_\xi$  і  $\varphi_\zeta$ :

$$\varphi_\xi(t_1) = \exp \{i t_1' \mu_1 - t_1' \Sigma_{11} t_1\}, \quad \varphi_\zeta(t_2) = \exp \{i t_2' \mu_2 - t_2' \Sigma_{22} t_2\}, \quad (3.20)$$

$$\varphi_\gamma(t) = \varphi_\xi(t_1) \cdot \varphi_\zeta(t_2), \quad (3.21)$$

де  $t = (t_1, t_2, \dots, t_q, t_{q+1}, \dots, t_{q+r})$ ,  $t_1 = (t_1, t_2, \dots, t_q)$ ,  $t_2 = (t_{q+1}, \dots, t_{q+r})$  (див. властивість 5 багатовимірних х.ф.).

**3.14 Наслідок.** Для того, щоб компоненти випадкового вектора, розподіленого за  $N(\mu, \Sigma)$ , були взаємно незалежними один від одного, необхідно і достатньо,

щоб матриця  $\Sigma$  була діагональною (тобто щоб всі елементи матриці  $\Sigma$ , що не розташовані на головній діагоналі, дорівнювали 0).

**3.15 Наслідок.** Нехай вектор  $\gamma = \begin{pmatrix} \xi \\ \varsigma \end{pmatrix}$ , де  $\xi$  та  $\varsigma$  —  $q$ - вимірний та  $r$ - вимірний вектори відповідно, має нормальний розподіл  $N_{q+r}(\mu, \Sigma)$ , причому  $M\xi = \mu_1$ ,  $M\varsigma = \mu_2$ ,  $\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$  (див. (3.17), (3.18)). Тоді вектор

$$1. \theta = \begin{pmatrix} \xi + T\varsigma \\ \varsigma \end{pmatrix}, \quad (3.23)$$

де  $T = T_{q \times r}$  — деяка матриця, має нормальний розподіл і знайти його математичне сподівання  $M\theta$  та матрицю коваріацій  $\Psi$ .

**3.16 Лема.** Нехай вектор  $\gamma = \begin{pmatrix} \xi \\ \varsigma \end{pmatrix}$ , де  $\xi$  та  $\varsigma$  —  $q$ - вимірний та  $r$ - вимірний

вектори відповідно, має нормальний розподіл  $N_{q+r}(\mu, \Sigma)$ ,  $\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}$  (див.

рівність (3.18)), причому  $\Sigma_{22}$  являється невідродженою матрицею. Покладемо  $\theta$

$$= \begin{pmatrix} \xi + T\varsigma \\ \varsigma \end{pmatrix},$$

де

$$T = -\Sigma_{12}\Sigma_{22}^{-1}. \quad (3.29)$$

Тоді вектор  $\theta$  має  $(q+r)$ -вимірний нормальний розподіл з параметрами

$$M\theta = \begin{pmatrix} \mu_1 + T\mu_2 \\ \mu_2 \end{pmatrix}, \quad \text{де } \mu_1 = M\xi, \mu_2 = M\varsigma, \quad (3.30)$$

i

$$D \theta = \begin{pmatrix} \Psi_{11} & \Psi_{12} \\ \Psi_{21} & \Psi_{22} \end{pmatrix}, \quad (3.31)$$

де

$$\Psi_{11} = \Sigma_{11} - \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21}, \quad \Psi_{12} = \mathbf{O}_{q \times r}, \quad \Psi_{21} (= \Psi_{12}') = \mathbf{O}_{r \times q}, \quad \Psi_{22} = \Sigma_{22}. \quad (3.32)$$

Зокрема, нормальні випадкові вектори  $\xi + T\zeta$  і  $\zeta$  є взаємно незалежними і мають параметри  $(\mu_1 + T\mu_2, \Psi_{11})$  та  $(\mu_2, \Psi_{22})$  відповідно один одному.

Знову розглядаючи вектор  $\gamma = \begin{pmatrix} \xi \\ \zeta \end{pmatrix}$ , що має сумісний нормальний розподіл,

знайдемо умовний розподіл вектора  $\xi$  при даному  $\eta$ .

**3.17 Теорема.** Нехай  $\gamma = \begin{pmatrix} \xi \\ \zeta \end{pmatrix} \in N_{q+r}(\mu, \Sigma)$  і має місце наступне розбиття вектора  $\mu$  і матриці  $\Sigma$  (див. рівності (3.17) і (3.18)):

$$\mu = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix}, \quad \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}, \quad (3.33)$$

де  $\mu_1 = M\xi$ ,  $\mu_2 = M\zeta$ ,  $\Sigma_{11} = D\xi$ ,  $\Sigma_{22} = D\zeta$ , причому, за умовою,  $\Sigma_{22}$  є невиродженою матрицею. Тоді умовний розподіл вектора  $\xi$  при  $\zeta = y$  є нормальним з параметрами

$$M(\xi \mid \zeta = y) = v(y) = \mu_1 + \Sigma_{12} \Sigma_{22}^{-1} (y - \mu_2), \quad (3.34)$$

$$D(\xi \mid \zeta = y) = \Sigma_{11.2} = \Sigma_{11} - \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21}. \quad (3.35)$$

Зокрема, характеристична функція  $\varphi_{\xi|\zeta}(\mathbf{x} | \mathbf{y})$  цього розподілу визначається рівністю:

$$\varphi_{\xi|\zeta}(\mathbf{x} | \mathbf{y}) = \exp \{ i \mathbf{x}' \mathbf{v}(\mathbf{y}) - \mathbf{x}' \boldsymbol{\Sigma}_{11.2} \mathbf{x} / 2 \}. \quad (3.36)$$

**3.18 Зауваження стосовно умовних щільностей нормально розподілених випадкових векторів.** У попередніх розділах ми вже зазначили, що якщо характеристична функція (для нормального) випадкового вектора представлена у вигляді (3.4), то при умові, що матриця коваріації  $\boldsymbol{\Sigma}$  не є виродженою, ймовірнісний розподіл має щільність, що визначається за формулою (3.8) Характеристична функція умовного розподілу (3.36), з очевидністю, має вигляд (3.4) з параметрами  $\boldsymbol{\mu} = \mathbf{v}(\mathbf{y})$  та  $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}_{11.2}$ , які визначаються в рівняннях (3.34), (3.35). Таким чином, умовою, яка є однаково необхідною і достатньою для існування щільності даного умовного розподілу, є невиродженість матриці  $\boldsymbol{\Sigma}_{11.2}$ . Позначивши цю щільність як  $f_{\xi|\zeta}(\mathbf{x}, \mathbf{y})$ , з рівності (3.8) ми отримуємо:

$$f_{\xi|\zeta}(\mathbf{x}, \mathbf{y}) = (2\pi)^{-q/2} (\det \boldsymbol{\Sigma}_{11.2})^{-1/2} e^{-\frac{1}{2}(\mathbf{x}-\mathbf{v}(\mathbf{y}))' \boldsymbol{\Sigma}_{11.2}^{-1} (\mathbf{x}-\mathbf{v}(\mathbf{y}))},$$

де параметри  $\mathbf{v}(\mathbf{y})$  та  $\boldsymbol{\Sigma}_{11.2}$  визначені рівностями (3.33) — (3.35).

### Про деякі способи перевірки нормальності векторних даних

Нехай маємо вибірку векторних даних  $\mathbf{X} = (x_1, x_2, \dots, x_n)$ , де  $x_j = \begin{pmatrix} x_{1j} \\ x_{2j} \\ \dots \\ x_{mj} \end{pmatrix}$  (так

що кожне  $x_j$  — вектор з  $m$  дійсними координатами). Гіпотеза  $H_0$ :

теоретичним розподілом даної вибірки можна вважати  $m$ - вимірний нормальний розподіл. Іншими словами, наявні вибіркові значення є реалізаціями векторної випадкової величини ( $m$ - вимірного випадкового вектора), що має  $m$ - вимірний нормальний розподіл.

Згідно з загальними принципами перевірки статистичних гіпотез для перевірки  $H_0$  слід обрати так звану множину прийняття гіпотези, критичну множину і статистику критерію (див. , наприклад, [33,34] ).

Виділимо ще дві роботи по даній тематиці, в яких використовуються явні аналітичні вирази статистики перевірки. Тест Мардіа [34] заснований на багатовимірних розширеннях мір асиметрії та ексцесу. В цитованій роботі використовуються наступні статистики: для вибірки  $\{\mathbf{x}_1, \dots, \mathbf{x}_n\}$   $k$ -вимірних векторів обчислюються значення виразів

$$\widehat{\Sigma} = \frac{1}{n} \sum_{j=1}^n (\mathbf{x}_j - \bar{\mathbf{x}})(\mathbf{x}_j - \bar{\mathbf{x}})' \quad (3.37)$$

$$A = \sum_{i=1}^n \sum_{j=1}^n [(\mathbf{x}_i - \bar{\mathbf{x}})' \widehat{\Sigma}^{-1} (\mathbf{x}_j - \bar{\mathbf{x}})]^3 \quad (3.38)$$

$$B = \sqrt{\frac{n}{8k(k+2)}} \left\{ \frac{1}{n} \sum_{i=1}^n [(\mathbf{x}_i - \bar{\mathbf{x}})' \widehat{\Sigma}^{-1} (\mathbf{x}_j - \bar{\mathbf{x}})]^2 - k(k+2) \right\} \quad (3.39)$$

Тут

$$\bar{\mathbf{x}} = \frac{1}{n} \sum_{i=1}^n \mathbf{x}_i,$$

$$\widehat{\Sigma} = \frac{1}{n} \sum_{i=1}^n (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})'$$

— оцінки теоретичного математичного сподівання і коваріаційної матриці розподілу, відповідно.

Встановлено, що за справедливостю нульової гіпотези статистика  $A$  наближено має розподіл  $\chi^2$  (див. , наприклад, [35] ) з  $1/6 \cdot k(k+1)(k+2)$  степенями свободи, а  $B$  апроксимується стандартним нормальним законом  $N(0,1)$ .

Відомо, що статистика ексцесу  $B$  Мардіа скошена і дуже повільно збігається до граничного нормального розподілу. Для вибірок середнього розміру ( $50 \leq n < 400$ ) параметри асимптотичного розподілу статистики ексцесу модифіковані в



роботі [36] Для невеликої вибірки використовуються емпіричні критичні значення ( $n < 50$ ). Таблиці критичних значень для обох статистик А, В наводяться в [36] для  $k = 2, 3, 4$ .

Тести Мардіа є афінно інваріантними, але у загальному випадку не консистентними. Наприклад, багатоваріантний тест асиметрії не є консистентним при симетричній ненормальній альтернативі [36].

### **Метод ПССВ (перехід до сукупності скалярних вибірок)**

Даний метод не використовує оціночних значень коваріаційних матриць і обчислення обернених матриць великої розмірності. Основна ідея зазначеного методу полягає у заміні роботи з характеристиками векторної вибірки дослідженням характеристик сукупності деяких скалярних вибірок. Ці вибірки є лінійними перетвореннями вихідної векторної вибірки. Згідно з теоремою 3.8, перевірку нормальності вихідної вибірки можна замінити перевіркою нормальності *всіх* таких перетворень. Практично мова може йти про перевірку нормальності *як завгодно великої кількості* таких перетворень. Деталі щодо зазначеного методу можна знайти в роботі [37].

### **Функції програмного забезпечення**

Програмне забезпечення, що розроблено у даній роботі, виконує перевірку гіпотези про співпадіння коваріаційних матриць кількох груп векторних даних однакової розмірності. Обчислюються, значення статистик критерію, котрі введені в даному розділі в підрозділі «**Деякі способи перевірки однорідності даних**», п.2). Умова коректності застосування програми — нормальність вибірових даних або принаймні близькість до нормального розподілу.

Програма розроблена в середовищі **Delphi Embarkadero RAD Studio**

### **Фрагменти програмного коду**

```
unit MainUnit;  
interface
```

uses

Windows, Messages, SysUtils, Variants, Classes, Graphics, Controls, Forms,  
Dialogs, StdCtrls, ComCtrls, Buttons, Menus, ExtCtrls, Grids, NewProjectUnit,  
Matrices, fDistr;

type

```
TMainForm = class(TForm)
  TopPanel: TPanel;
  LoadButton: TBitBtn;
  SaveButton: TBitBtn;
  StatusBar: TStatusBar;
  RightPanel: TGroupBox;
  LeftPanel: TGroupBox;
  ExitButtonPanel: TPanel;
  InputPageControl: TPageControl;
```

begin

```
xf := TXLSFile.Create;
```

try

```
xf.OpenFile(FileName);
```

```
MainForm.MatrixCountEdit.Text := IntToStr(xf.Workbook.Sheets.Count);
```

```
matrix0.ColCount := 1;
```

```
matrix0.RowCount := 1;
```

```
for i := 0 to xf.Workbook.Sheets[0].Cells.Count - 1 do
```

```
begin
```

```
  C := xf.Workbook.Sheets[0].Cells.Item[i];
```

```
  if (matrix0.ColCount < C.Col + 1) and (C.Value <> "")
```

```
    then matrix0.ColCount := C.Col + 1;
```

```
  if (matrix0.RowCount < C.Row + 1) and (C.Value <> "")
```

```
    then matrix0.RowCount := C.Row + 1;
```

```
  if (C.Value <> "") and (VarIsNumeric(C.Value)) then matrix0.Cells[C.Col, C.Row] :=
```

```
FloatToStr(C.Value);
```

```
  if (C.Value <> "") and (not VarIsNumeric(C.Value)) then matrix0.Cells[C.Col,  
C.Row] := C.Value;
```

```
end;
```

```
if xf.Workbook.Sheets.Count > 1 then
```

```

begin
  matrix1.ColCount := 1;
  matrix1.RowCount := 1;
  for i := 0 to xf.Workbook.Sheets[1].Cells.Count - 1 do
  begin
    C := xf.Workbook.Sheets[1].Cells.Item[i];
    if (matrix1.ColCount < C.Col + 1) and (C.Value <> "")
      then matrix1.ColCount := C.Col + 1;
    if (matrix1.RowCount < C.Row + 1) and (C.Value <> "")
      then matrix1.RowCount := C.Row + 1;

    if (C.Value <> "") and (VarIsNumeric(C.Value)) then matrix1.Cells[C.Col, C.Row]
:= FloatToStr(C.Value);
    if (C.Value <> "") and (not VarIsNumeric(C.Value)) then matrix1.Cells[C.Col,
C.Row] := C.Value;
    end;
  end;
procedure TMainForm.ShowResults(count: Byte);
begin
  if count > 0 then
  begin
    OutputPage0.TabVisible := True;
    WriteVisulaMatrix(OutputMatrix0, ResultMatrix[0]);
    OutputMatrix0.Visible := True;
  end;
  if count > 1 then
  begin
    b := 0;
    m1 := 0;
    m2 := 0;
    Res1 := 0;
    Res2 := 0;

    MatrixCount := StrToInt(MatrixCountEdit.Text);

```

```

SetLength(InpMatrix, MatrixCount);
SetLength(YMatrix, MatrixCount);
SetLength(ResultMatrix, MatrixCount + 1);
if InputPage0.TabVisible then InpMatrix[0] := ReadVisulaMatrix(InputMatrix0);
if InputPage1.TabVisible then InpMatrix[1] := ReadVisulaMatrix(InputMatrix1);
if InputPage2.TabVisible then InpMatrix[2] := ReadVisulaMatrix(InputMatrix2);
if InputPage3.TabVisible then InpMatrix[3] := ReadVisulaMatrix(InputMatrix3);
if InputPage4.TabVisible then InpMatrix[4] := ReadVisulaMatrix(InputMatrix4);
if InputPage5.TabVisible then InpMatrix[5] := ReadVisulaMatrix(InputMatrix5);

ResultMatrix[i] := Mul(YMatrix[i], Transpose(YMatrix[i]));

for i := 0 to MatrixCount - 1 do
begin
for m := 0 to ResultMatrix[i].M - 1 do
begin
for n := 0 to ResultMatrix[i].N - 1 do
begin
ResultMatrix[i].Data[m, n] := ResultMatrix[i].Data[m, n] / InpMatrix[i].N;
end;
end;
end;
end;

```

### **Керівництво користувача:**

1. Завантажте програму, запустивши файл App.exe
2. Завантажте готовий приклад, вибравши "завантажити" або створіть новий приклад. В останньому випадку задайте кількість матриць, кількість рядків (висота матриці) і кількість стовпців (ширина матриці) для кожної матриці.

***Зауваження: всі матриці повинні мати однакову кількість рядків (причина: порівняння виконується для векторів однакової розмірності)***

3. Заповніть порожні клітини матриць відповідними вибілковими значеннями.

3. Натисніть "розрахувати"

*У правій частині екрану з'явиться нова матриця. Це та сама матриця  $S_U$ , що фігурує у співвідношеннях підрозділу 2) «Перевірка гіпотези про рівність кількох коваріаційних матриць» даного розділу) З'явиться також напис «Дані однорідні» або «Дані неоднорідні»*

4. Отримайте та передивіться результати роботи програми, що з'являться в мемо полі. Надайте змістової інтерпретації одержаній відповіді.

Приклад роботи програми

Матриця 0

|   |   |   |   |   |   |
|---|---|---|---|---|---|
| 2 | 0 | 1 | 4 | 2 | 1 |
| 3 | 2 | 1 | 4 | 3 | 0 |
| 3 | 3 | 2 | 1 | 0 | 1 |

Матриця 1

|   |   |   |   |   |   |
|---|---|---|---|---|---|
| 2 | 2 | 0 | 3 | 2 | 1 |
| 1 | 2 | 1 | 4 | 2 | 2 |
| 0 | 2 | 1 | 3 | 5 | 0 |

Наведені дані (матриця 0 і матриця 1) представляють відомості щодо кількості продажів трьох видів товарів в двох різних пунктах реалізації за один тиждень. Чи можна вважати, що розподіли продажів по днях тижня однакові для цих двох пунктів? Наступний програмний інтерфейс відповідає розв'язку вказаної задачі. Відповідь позитивна.

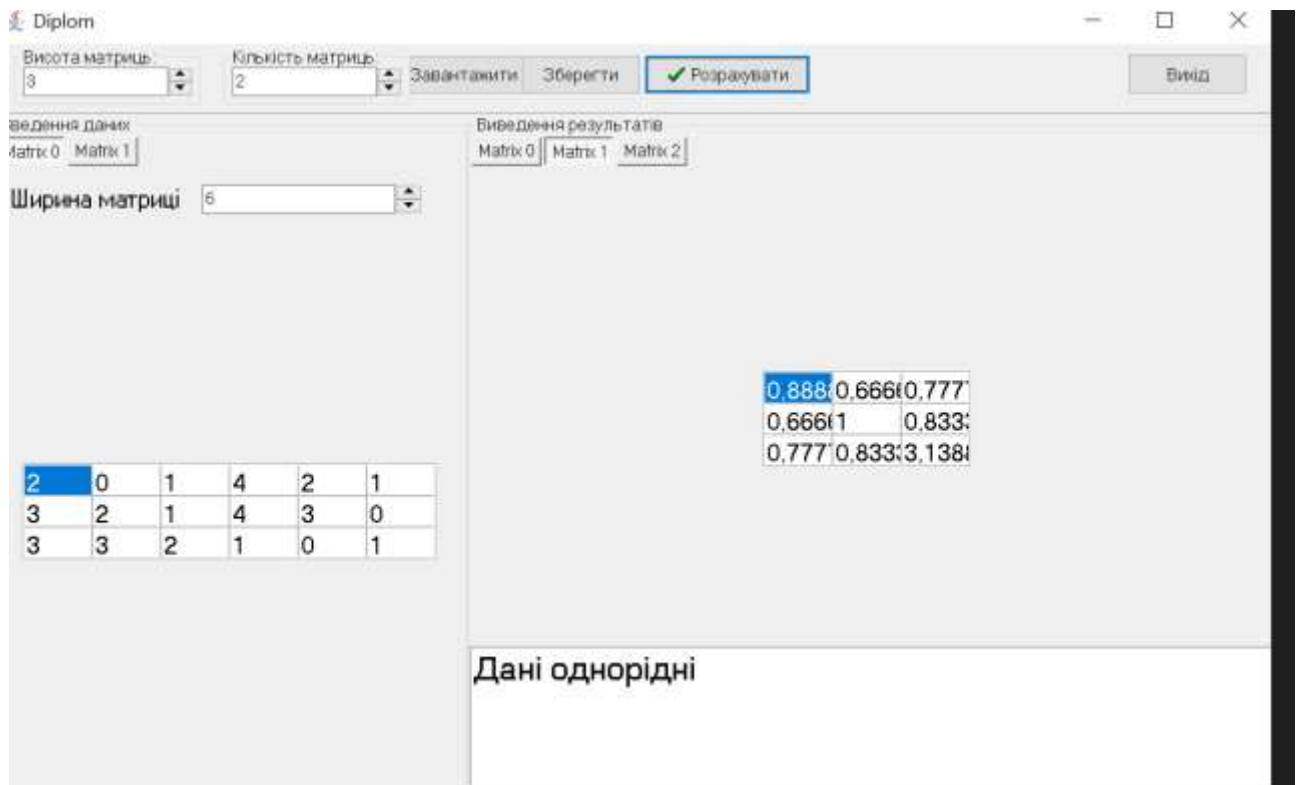


Рисунок 3.1. Програмний інтерфейс

### Висновки до 3 розділу

Розроблено програмне забезпечення, що реалізує методи перевірки однорідності вибірок багатовимірних статистичних даних. Показником однорідності за означенням, прийнятим у даній роботі, є позитивний результат вирішення задачі перевірки гіпотези про співпадіння теоретичних коваріаційних матриць даної системи багатовимірних вибірок. Виконується також обчислення критичної точки розподілу Фішера  $F_{v_1, v_2}$  і формування висновку щодо гіпотези про співпадіння коваріаційних матриць досліджуваних векторів.

Обчислення належних імовірнісних характеристик розподілів, що виконують роль індикаторів перевірки, (квантилів розподілів Фішера) виконується самою програмою. Як наслідок, звертання до спеціальних статистичних програм або таблиць не потрібне. Дане програмне забезпечення може бути застосовано для розв'язання часто виникаючих на практиці задач порівняння результатів різних технологічних процесів, опитувань різних верств населення, соціальних стандартів.

## **Висновки**

В даному дипломному проекті ми розглянули важливість та використання методів багатовимірного статистичного аналізу в економічних процесах в умовах нестабільності, неповності та часткової інформації. Ми докладно розглянули базові поняття бізнес-аналітики та представили аналіз ринку програмного забезпечення для багатовимірного статистичного аналізу даних.

Також були висвітлені різноманітні аспекти використання цих методів у бізнес-аналітиці, включаючи кореляційний та регресійний аналіз, класифікацію, вибір репрезентативних ознак та зменшення розмірності даних. Також ми зазначили важливість програмного забезпечення для вирішення завдань перевірки однорідності вибірок багатовимірних статистичних даних, що може використовуватися для порівняння різних технологічних процесів та соціальних стандартів.



## СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Ермантраут Е. Р. Статистичний аналіз агрономічних дослідних даних в пакеті Statistica-6 / Е.Р. Ермантраут, О.І. Присяжнюк, І.Л. Шевченко.— К. : Поліграф Консалтинг, 2007. – 55 с.
2. Караєва Н. В. Еколого-економічна оптимізація виробництва: інформаційна підтримка прийняття рішень: конспект лекцій. –К.:НТУУ «КПІ», 2016. – 115 с.
3. Левченко Л.О., Кілянчук О.П., Повханич О.Ю. Огляд програмних продуктів фінансово-економічного аналізу діяльності енергопідприємств [Електронний ресурс]. – Режим доступу: <http://urss.knuba.edu.ua/files/zbirnyk-8/121-127.pdf> (дата звернення 01.10.2023). – Назва з екрану.
4. Майборода Р.Є., Сугакова О.В. Статистичний аналіз даних за допомогою пакету STATISTICA [Електронний ресурс]. – Режим доступу: <http://matphys.rpd.univ.kiev.ua/downloads/courses/mmatstat/StatAn.doc> (дата звернення 01.10.2023). – Назва з екрану.
5. Присяжнюк О. І. Комплексна оцінка сучасних гібридів цукрових буряків/ О.І.Присяжнюк, Т.Д.Сонєць, О.Ю.Половинчук, І.І.Коровко// Наукові праці інституту біоенергетичних культур і цукрових буряків.- 2016.- N 24. - С.18-27.
6. Програми для математичної і статистичної обробки даних [Електронний ресурс]. – Режим доступу: <http://chem-bio.com.ua/aspirant/grant/item/> (дата звернення 01.10.2023). – Назва з екрану.
7. Програми для наукових розрахунків [Електронний ресурс]. – Режим доступу: <http://softlist.com.ua/catalog/programmy-dlya-nauchnykh-raschetov/> (дата звернення 01.10.2023). – Назва з екрану
8. Єгоршин О. О. Методи багатовимірного статистичного аналізу : навч. посіб. / О. О. Єгоршин, А. М. Зосімов, В. С. Пономаренко. – Київ : ІЗМН, 1998.– 208 с.

9. Клебанова Т. С. Механізм та моделі управління кризовими ситуаціями на підприємствах житлово-комунального комплексу / Т. С. Клебанова, М. О. Кизим, Ю. І. Мізік. – Харків : ВД "ІНЖЕК", 2011. – 178 с.
10. Клебанова Т. С. Нечітка логіка та нейронні мережі в управлінні підприємством / Т. С. Клебанова, Л. О. Чаговець, О. В. Панасенко. – Харків : ВД "ІНЖЕК", 2011. – 239 с.
11. Куліков П. М. Економіко-математичне моделювання фінансово-го стану підприємства / П. М. Куліков. – Харків : ІНЖЕК, 2009. – 151 с.
12. Математичні методи в сучасних економічних дослідженнях / Л. М. Малярець, О. Г. Тижненко, О. О. Єгоршин та ін.; за ред. Л. М. Малярець. – Харків : Вид. ХНЕУ, 2011. – 270 с.
13. Математичні методи і моделі ринкової економіки / Т. С. Клебанова, М. О. Кизим, О. І. Черняк та ін.; за ред. Т. С. Клебанової. – Харків : ВД "ІНЖЕК", 2010. – 454 с.
14. Моделі і методи соціально-економічного прогнозування : підручник / В. М. Геєць, Т. С. Клебанова, О. І. Черняк та ін.; за ред. В. М. Гейця. – Харків : ВД "ІНЖЕК", 2005. – 396 с.
15. Моделювання системних характеристик в економіці / О. М. Сінчук, Т. М. Берідзе, В. В. Кононенко та ін.; за ред. О. М. Сінчук. – Кременчук : ПП Щербатих О. В., 2009. – 211 с.
16. Практикум з навчальної дисципліни "Багатомірний статистичний аналіз" для студ. спец. "Прикладна економіка" ден. форми навч. / Т. С. Клебанова, Л. С. Гур'янова, О. А. Сергієнко. – Харків : ХНЕУ, 2011. – 74 с.
17. Прогнозування соціально-економічних процесів / Т. С. Клебанова, В. А. Курзенев., В. М. Наумов та ін.; за ред. Т. С. Клебанової. – Харків : ХНЕУ ім. С. Кузнеця, 2015. – 654 с.

18. Годун В. М. Інформаційні системи і технології в статистиці / В. М. Годун, Н. С. Орленко, М. А. Сендзюк [Електронний ресурс] – Режим доступу : <http://library.if.ua/book/80/5668.html>.
19. Практичний досвід інформаційно-аналітичної підтримки процедур розробки і прийняття управлінських рішень [Електронний ресурс]. – Режим доступу : [http://www.ecsor.com.ua/files/conf\\_report\\_2\\_ukr.pdf](http://www.ecsor.com.ua/files/conf_report_2_ukr.pdf).
20. Birch M. W. A new proof of the Pearson–Fisher theorem / M. W. Birch // *Ann. Math. Statist.* – 1964. – V. 35. – P. 817.
21. Eisenberger J. Systematic statistics used for data compression in space telemetry / J. Eisenberger, E.C. Posner // *J. Amer. Statist. Ass.* 60. – 1965. – Pp. 97–133.
22. Gupta S.S. Estimation of the parameters of the logistic distribution / S.S. Gupta, M. Gnanadesikan // *Biometrika*, 53, 1966. – Pp. 565–570.
23. Hampel F. R. The influence curve and its role in robust estimation / F. R. Hampel // *J. Amer. Statist. Ass.* – 1974. – V. 69. – № 346. – Pp. 383–393.
24. Ogawa J. Contributions to the theory of systematic statistics / J.Ogawa // *I. Osaka Math. J.* 3, 1951. – Pp. 175–213.
25. Saleh A.K.M.J. Asymptotic optimum quantiles for the estimation of the parameters of the negative exponential distribution / A.K.M.J. Saleh, M.M. Ali // *Ann. Math. Statist.* 37, 1966. – Pp. 143–151.
26. 830-1998 - IEEE Recommended Practice for Software Requirements Specifications [Electronic resource]. - Mode of access: <http://standards.ieee.org/findstds/standard/830-1998.html> (Accessed 10 Oct 2023). – Title from the screen.
27. Mardia, R., Zemroch, P. (1978). *Tables of the F- and Related Distributions With Algorithms*. London: Academic Press, 250.
28. Albert A. Bennett and James O. Berger. *Statistical Tables for the Social, Biological, and Physical Sciences* by: Springer, 2021, 632 p
29. Tukey, J.W. (1977). *Exploratory Data Analysis*. Massachusetts: Addison Wesley Publishing Company Reading, 693. <https://www.amazon.com/Exploratory-Data-Analysis-John-Tukey/dp/0201076160>

30. Johnson, R.A., Wichern. D.W. Applied Multivariate Statistical Analysis (2017). Washington: Pearson, 773. <https://www.ojp.gov/ncjrs/virtual-library/abstracts/factor-analysis-statistical-methods-and-practical-issues>
31. Краснитський С.М., Щербань В.Ю. Краснитський М.С., Заржицький Є.В. та ін. Векторні випадкові величини і випадкові процеси. — К.: Бумсервіс, 2018. — 191 с.
32. Kim, J.O., Mueller, C.W. Factor Analysis: Statistical Methods and Practical Issues (2011). London: Sage Publications, 300. <https://www.ojp.gov/ncjrs/virtual-library/abstracts/factor-analysis-statistical-methods-and-practical-issues>
33. Краснитський С.М., Щербань В.Ю., Резанова В.Г. Основні поняття математичної статистики. — К.: КНУТД, 2016. — 96 с.
34. *Mardia, K. V. (1970). Measures of multivariate skewness and kurtosis with applications // Biometrika. 57 (3): 519–530.*
35. Булдигін В.В., Дем'яненко О.О. Асимптотичні властивості емпіричної корелограми гауссових векторних полів // Доп. АН України. — 1997. — №1. — С. 45 — 49
36. RENCHER, A. C. Methods of Multivariate Analysis A JOHN WILEY & SONS, INC. PUBLICATION 2002, 695 p.
37. Заржицький Є.В., Краснитський М.С. Комп'ютерна програма для перевірки нормальності закону розподілу векторної вибірки // Комплексна дипломна робота на здобуття ступеня магістра комп'ютерних наук — Київ, КНУТД, 2005.