МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
МІЖНАРОДНИЙ ГУМАНІТАРНИЙ УНІВЕРСИТЕТ
Кафедри комп’ютерних наук та інформаційних технологій

Григор’єва Т.І., Соловська І.М.

АНАЛІЗ ДАНИХ ТА КОМП’ЮТЕРНА СТАТИСТИКА

Методичні рекомендації для самостійної роботи здобувачів

Одеса 2024

Затверджено Вченою Радою Міжнародного гуманітарного університету (протокол № 13 від 16.08.2024 р.)

Григор'єва Т.І., Соловська І.М.
Аналіз даних та комп'ютерна статистика: методичні рекомендації для самостійної роботи здобувачів [Електронне видання]. / Григор'єва Т.І., Соловська І.М. Кафедри комп'ютерних наук та інформаційних технологій Міжнародного гуманітарного університету. Одеса, 2024. – 50 с.

Методичні рекомендації з курсу «**Аналіз даних та комп'ютерна статистика**» розроблено відповідно до навчального плану, вони складаються з навчальної програми курсу, методичних рекомендацій з проведення практичних занять і завдань для самостійної роботи здобувачів, списку рекомендованої літератури. Матеріали призначено для студентів Міжнародного гуманітарного університету, які в магістратурі вивчають галузь знань - 11 Математика та статистика.

Вивчення дисципліни «**Аналіз даних та комп'ютерна статистика**» сприятиме залученню здобувачів до науково-дослідницької діяльності та підготовки ними публікацій, кваліфікаційних й інших наукових робіт.

ОПИС НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

<table>
<tr><th rowspan="2">Найменування показників</th><th rowspan="2">Галузь знань, напрям підготовки, освітньо-кваліфікаційний рівень</th><th>Характеристика навчальної дисципліни</th></tr>
<tr><th>денна форма навчання</th></tr>
<tr><td rowspan="10">Кількість кредитів – 4

Загальна кількість годин – 120</td><td>Галузь - 11 Математика та статистика</td><td>обов’язкова</td></tr>
<tr><td rowspan="9">Спеціальність – 112 Статистика</td><td>Рік підготовки:</td></tr>
<tr><td>1-й</td></tr>
<tr><td>Семестр</td></tr>
<tr><td>1-й</td></tr>
<tr><td>Лекції</td></tr>
<tr><td rowspan="8">Мова навчання – українська</td><td rowspan="8">Рівень вищої освіти – другий (магістерський) рівень</td><td>14 год.</td></tr>
<tr><td>Практичні, семінарські</td></tr>
<tr><td>26 год.</td></tr>
<tr><td>Лабораторні</td></tr>
<tr><td>-</td></tr>
<tr><td>Самостійна робота та індивідуальні завдання</td></tr>
<tr><td>80 год.</td></tr>
<tr><td>Вид контролю:</td></tr>
<tr><td>екзамен</td></tr>
</table>

Дисципліна **"Аналіз даних та комп’ютерна статистика"** вивчає сучасні методи збору, обробки, аналізу та інтерпретації даних з використанням комп'ютерних технологій і статистичних методів, які забезпечують студентів знаннями та практичними навичками, необхідними для вирішення широкого спектру задач в різних галузях на основі сучасних методів аналізу даних та статистики.

Метою Метою викладення дисципліни "Аналіз даних та комп’ютерна статистика" є надання студентам навичок у використанні комп'ютерних засобів для проведення статистичних досліджень, аналізу великих обсягів даних і побудови математичних моделей, які використовуються для прогнозування та ухвалення рішень.

Передумови для вивчення дисципліни є знання і вміння, отримані студентом при вивченні навчальних дисциплін бакалаврської підготовки.

ЗАПЛАНОВАНІ РЕЗУЛЬТАТИ НАВЧАННЯ ЗА НАВЧАЛЬНОЮ ДИСЦИПЛІНОЮ

Знання:

- Знання основних принципів і методів описової та індуктивної статистики.
- Знання основ ймовірнісних розподілів та їх застосувань у статистичному аналізі.
- Розуміння методів перевірки статистичних гіпотез.
- Знання різних статистичних моделей, включаючи регресійний аналіз, кореляційний аналіз, дисперсійний аналіз, кластерний аналіз, факторний аналіз та дискримінантний аналіз.

Уміння:

- Уміння проводити попередню обробку даних, включаючи очищення, нормалізацію та трансформацію даних.
- Уміння застосовувати різні статистичні методи для аналізу даних, включаючи описову статистику, кореляційний і регресійний аналіз, кластерний і факторний аналіз.
- Уміння інтерпретувати результати статистичних аналізів та робити на їх основі обґрунтовані висновки.
- Уміння застосовувати непараметричні методи в аналізі даних.
- Уміння створювати інформативні графіки, діаграми та інші візуальні представлення даних.
- Уміння використовувати програмні пакети для проведення статистичних досліджень

Навички:

- Навички критичного мислення при виборі методів аналізу та інтерпретації результатів.
- Навички аналізу ризиків та невизначеностей при роботі з даними.
- Навички роботи з великими обсягами даних та їх обробки.
- Навички побудови та тестування статистичних моделей.
- Навички комунікації результатів аналізу, як в усній, так і в письмовій формі.
- Навички роботи в команді та спільної роботи над проектами аналізу даних.

ПРОГРАМА НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

Тема 1. Основні етапи статистичного аналізу.

Задачі, які розв'язують методами прикладної статистики. Мода, медіана, середнє та варіація. Інтервальні оцінки. Основні аспекти інтервальних оцінок. Реалізація в статистичних пакетах.

Тема 2. Перевірка гіпотез.

Типи задач, де застосовується критерій Пірсона.Типи задач, де застосовується критерій Стьюдента. U-критерій Манна-Уїтні.

Тема 3. Дисперсійний аналіз.

Основні типи дисперсійного аналізу. Дисперсійний аналіз і його реалізація в статистичних пакетах.

Тема 4. Кореляційний аналіз.

Основні методи кореляційного аналізу. Коефіцієнт кореляції Пірсона. Коефіцієнт кореляції Спірмена. Коефіцієнт кореляції Кендалла. Множинний кореляційний аналіз. Реалізація кореляційного аналізу в статистичних пакетах.

Тема 5. Регресійний аналіз.

Основні етапи регресійного аналізу. Реалізація регресійного аналізу в статистичних пакетах. Логістична регресія в статистичних пакетах

Тема 6. Багатовимірний статистичний аналіз.

Кластерний аналіз і його реалізація в статистичних пакетах. Дискримінантний аналіз і його реалізація в статистичних пакетах. Факторний аналіз і його реалізація в статистичних пакетах.

СТРУКТУРА НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

Назви змістових модулів і тем	Кількість годин									
	Денна форма					Заочна форма				
	Усього	у тому числі				Усього	у тому числі			
		Лекц.	Прак.	Лаб.	Сам. роб.		Лекц.	Прак.	Лаб.	Сам. роб.
Тема 1. Основні етапи статистичного аналізу	16	2	4		10	16		2		14
Тема 2. Перевірка гіпотез	18	4	4		10	18	2	2		14
Тема 3. Дисперсійний аналіз	26	2	4		20	26	2	2		22
Тема 4. Кореляційний аналіз	26	2	4		20	26	2	2		22
Тема 5. Регресійний аналіз	18	2	6		10	18	2	2		14
Тема 6. Багатовимірний статистичний аналіз	16	2	4		10	16		2		14
Усього годин	**120**	**14**	**26**	**-**	**80**	**120**	**8**	**12**	**-**	**100**
ПІДСУМКОВИЙ КОНТРОЛЬ – ЕКЗАМЕН										

ПИТАННЯ ДО ПРАКТИЧНИХ ЗАНЯТЬ

Тема 1. Основні етапи статистичного аналізу.

Задачі, які розв’язують методами прикладної статистики. Мода, медіана, середнє та варіація. Інтервальні оцінки. **Основні аспекти інтервальних оцінок.** Реалізація в статистичних пакетах.

Тема 2. Перевірка гіпотез.

Типи задач, де застосовується критерій Пірсона. Типи задач, де застосовується критерій Стьюдента. U-критерій Манна-Уїтні.

Тема 3. Дисперсійний аналіз.

Основні типи дисперсійного аналізу. Дисперсійний аналіз і його реалізація в статистичних пакетах.

Тема 4. Кореляційний аналіз.
Основні методи кореляційного аналізу. Коефіцієнт кореляції Пірсона. Коефіцієнт кореляції Спірмена. Коефіцієнт кореляції Кендалла. Множинний кореляційний аналіз. Реалізація кореляційного аналізу в статистичних пакетах.

Тема 5. Регресійний аналіз.

Основні етапи регресійного аналізу. Реалізація регресійного аналізу в статистичних пакетах. Логістична регресія в статистичних пакетах

Тема 6. Багатовимірний статистичний аналіз.

Кластерний аналіз і його реалізація в статистичних пакетах. Дискримінантний аналіз і його реалізація в статистичних пакетах. Факторний аналіз і його реалізація в статистичних пакетах.

САМОСТІЙНА РОБОТА

До самостійної роботи студентів включаються:

1. Знайомство з науковою та навчальною літературою відповідно зазначених у програмі тем.
2. Опрацювання лекційного матеріалу.
3. Підготовка до практичних занять.
4. Консультації з викладачем протягом семестру.
5. Самостійне опрацювання окремих питань навчальної дисципліни.
6. Підготовка до підсумкового контролю.

Тематика та питання до самостійної підготовки та індивідуальних завдань

Тема 1. Основні етапи статистичного аналізу.

Типові задачі прикладної статистики: прогнозування, класифікація, оцінка параметрів та перевірка гіпотез. Вибір рівня довіри та його вплив на інтервальну оцінку. Приклади застосування інтервальних оцінок у реальних дослідженнях. Порівняння інтервальних та точкових оцінок: переваги та обмеження.

Тема 2. Перевірка гіпотез.

Процес прийняття рішень: рівень значущості, критична область, p-значення. Приклади використання критерію Пірсона у соціології, маркетингу та біології. t-критерій Стьюдента для порівняння середніх значень. Застосування t-критерію для незалежних та залежних вибірок. Приклади використання критерію Стьюдента в економіці та медицині. Опис U-критерію Манна-Уїтні як непараметричного аналогу t-критерію. Типові задачі, де U-критерій використовується для порівняння двох незалежних груп. Приклади використання U-критерію Манна-Уїтні в психології та соціології.

Тема 3. Дисперсійний аналіз.

Поняття міжгрупової та внутрішньогрупової дисперсії. Приклад використання дисперсійного аналізу в економіці та біології. Модель однофакторного дисперсійного аналізу. Особливості двофакторного дисперсійного аналізу з та без взаємодії факторів. Приклад застосування двофакторного аналізу в соціологічних дослідженнях. Інтерпретація результатів та їх вплив на висновки дослідження.

Тема 4. Кореляційний аналіз.

Поняття кореляції та її види (лінійна, нелінійна, парна, множинна). Приклади задач, що розв’язуються методами кореляційного аналізу. Відмінності між коефіцієнтами кореляції Пірсона та Спірмена. Приклади використання коефіцієнта кореляції Спірмена для рангових даних. Опис коефіцієнта кореляції Кендалла та сфери його застосування. Приклади задач, що використовують цей коефіцієнт для аналізу. Приклад використання множинного кореляційного аналізу у маркетингових дослідженнях. Реалізація множинного кореляційного аналізу в статистичних пакетах.

Тема 5. Регресійний аналіз.

Постановка задачі регресійного аналізу та вибір моделі. Приклади застосування лінійної регресії в економіці та фінансах. Основні методи побудови нелінійних регресійних моделей. Застосування логістичної регресії для прогнозування бінарних результатів. Приклади використання логістичної регресії в медицині та соціології. Особливості побудови та інтерпретації множинної регресії.

Тема 6. Багатовимірний статистичний аналіз.

Визначення та сфери застосування багатовимірного статистичного аналізу. Основні методи аналізу багатовимірних даних: кластерний, дискримінантний, факторний аналіз. Основні методи кластерного аналізу: ієрархічний, k-means, методи на основі щільності. Особливості ієрархічного кластерного аналізу та побудова дендрограм. Переваги та недоліки методу k-means, варіанти покращення алгоритму. Основи дискримінантного аналізу та його використання для класифікації. Приклад задачі дискримінантного аналізу в медицині або фінансах. Приклади застосування факторного аналізу в психології, економіці та маркетингу. Спільні риси та відмінності між кластерним, дискримінантним та факторним аналізами.

ТЕМИ ДОПОВІДЕЙ

1. Типові задачі прикладної статистики: прогнозування, класифікація, оцінка параметрів та перевірка гіпотез. Приклади задач у бізнесі, економіці, медицині та соціології.
2. Порівняння та відмінності між середніми показниками: мода, медіана, середнє арифметичне.
3. Аналіз варіації та її роль у статистиці.
4. Приклади застосування інтервальних оцінок у реальних дослідженнях.
5. Оцінка параметрів розподілу за допомогою інтервальних оцінок.
6. Порівняння інтервальних та точкових оцінок: переваги та обмеження.
7. Розподіли з різними параметрами: нормальний, експоненціальний, рівномірний.
8. Приклади застосування аналізу розподілів у бізнесі та економіці.
9. Приклади використання критерію Стьюдента в економіці та медицині.
10. Приклади використання U-критерію Манна-Уїтні в психології та соціології.
11. Огляд основних непараметричних критеріїв (U-критерій Манна-Уїтні, критерій серій Вальда-Вольфовиця). Переваги та обмеження непараметричних методів.
12. Застосування статистичних критеріїв у бізнесі, медицині, соціології та маркетингу.
13. Приклад використання дисперсійного аналізу в економіці та біології.
14. Приклад застосування двофакторного аналізу в соціологічних дослідженнях.
15. Переваги та недоліки використання програмних пакетів для дисперсійного аналізу.

16. Приклад використання множинного кореляційного аналізу у маркетингових дослідженнях.
17. Приклади використання логістичної регресії в медицині та соціології.
18. Приклади застосування кластерного аналізу в маркетингових дослідженнях, соціології та біоінформатиці.
19. Приклад використання ієрархічного кластерного аналізу для сегментації ринків.
20. Приклади застосування факторного аналізу в психології, економіці та маркетингу.
21. Порівняльний аналіз кластерного, дискримінантного та факторного аналізу. Вибір методу залежно від специфіки даних та поставленої задачі.

ВИДИ ТА МЕТОДИ КОНТРОЛЮ

Робоча програма навчальної дисципліни передбачає наступні види та методи контролю:

Види контролю	Складові оцінювання
поточний контроль, який здійснюється у ході: проведення практичних занять, виконання індивідуального завдання; проведення консультацій та відпрацювань.	**50%**
підсумковий контроль, який здійснюється у ході проведення заліку.	**50%**

Методи діагностики знань (контролю)	фронтальне опитування; наукова доповідь, усне повідомлення, індивідуальне опитування, робота у групах, розв`язання задач і практичних завдань, екзамен

Питання до екзамену

1. Основні етапи статистичного аналізу.
2. Задачі, які розв’язують методами прикладної статистики.
3. Мода, медіана, середнє значення та варіація.
4. Інтервальні оцінки. Основні аспекти інтервальних оцінок. Реалізація в статистичних пакетах.
5. Перевірка гіпотез. Типи задач, де застосовується критерій Пірсона.
6. Типи задач, де застосовується критерій Стьюдента.
7. U-критерій Манна-Уїтні.
8. Дисперсійний аналіз. Основні типи дисперсійного аналізу.
9. Модель однофакторного дисперсійного аналізу.
10. Особливості двофакторного дисперсійного аналізу з та без взаємодії факторів. Приклад застосування двофакторного аналізу в соціологічних дослідженнях.
11. Дисперсійний аналіз і його реалізація в статистичних пакетах.

12. Кореляційний аналіз. Основні методи кореляційного аналізу.
13. Поняття кореляції та її види (лінійна, нелінійна, парна, множинна).
14. Коефіцієнт кореляції Пірсона.
15. Коефіцієнт кореляції Спірмена.
16. Відмінності між коефіцієнтами кореляції Пірсона та Спірмена.
17. Коефіцієнт кореляції Кендалла.
18. Опис коефіцієнта кореляції Кендалла та сфери його застосування. Приклади задач, що використовують цей коефіцієнт для аналізу.
19. Множинний кореляційний аналіз.
20. Реалізація кореляційного аналізу в статистичних пакетах.
21. Регресійний аналіз. Основні етапи регресійного аналізу.
22. Основні методи побудови нелінійних регресійних моделей.
23. Застосування логістичної регресії для прогнозування бінарних результатів. Приклади використання логістичної регресії в медицині та соціології.
24. Реалізація регресійного аналізу в статистичних пакетах.
25. Логістична регресія в статистичних пакетах.
26. Особливості побудови та інтерпретації множинної регресії.
27. Багатовимірний статистичний аналіз.
28. Кластерний аналіз і його реалізація в статистичних пакетах.
29. Дискримінантний аналіз і його реалізація в статистичних пакетах.
30. Факторний аналіз і його реалізація в статистичних пакетах.
31. Спільні риси та відмінності між кластерним, дискримінантним та факторним аналізами.

КРИТЕРІЇ ПІДСУМКОВОЇ ОЦІНКИ ЗНАНЬ СТУДЕНТІВ

(для іспиту / заліку)

Рівень знань оцінюється:

- «відмінно» / «зараховано» А - від 90 до 100 балів. Студент виявляє особливі творчі здібності, вміє самостійно знаходити та опрацьовувати необхідну інформацію, демонструє знання матеріалу, проводить узагальнення і висновки. Був присутній на лекціях та семінарських заняттях, під час яких давав вичерпні, обґрунтовані, теоретично і практично правильні відповіді, має конспект з виконаними завданнями до самостійної роботи, проявляє активність і творчість у науково-дослідній роботі;

- «добре» / «зараховано» B - від 82 до 89 балів. Студент володіє знаннями матеріалу, але допускає незначні помилки у формуванні термінів, категорій, проте за допомогою викладача швидко орієнтується і знаходить правильні відповіді. Був присутній на лекціях та семінарських заняттях, має конспект з виконаними завданнями до самостійної роботи, проявляє активність і творчість у науково-дослідній роботі;

- «добре» / «зараховано» C - від 74 до 81 балів. Студент відтворює значну частину теоретичного матеріалу, виявляє знання і розуміння основних положень, з допомогою викладача може аналізувати навчальний матеріал, але дає недостатньо обґрунтовані, невичерпні відповіді, допускає помилки. При цьому враховується наявність конспекту з виконаними завданнями до самостійної роботи та активність у науково-дослідній роботі;

- «задовільно» / «зараховано» D - від 64 до 73 балів. Студент був присутній не на всіх лекціях та семінарських заняттях, володіє навчальним матеріалом на середньому рівні, допускає помилки, серед яких є значна кількість суттєвих. При цьому враховується наявність конспекту з виконаними завданнями до самостійної роботи;

- «задовільно» / «зараховано» E - від 60 до 63 балів. Студент був присутній не на всіх лекціях та семінарських заняттях, володіє навчальним матеріалом на рівні, вищому за початковий, значну частину його відтворює на репродуктивному рівні, на всі запитання дає необґрунтовані, невичерпні відповіді, допускає помилки, має неповний конспект з завданнями до самостійної роботи.

- «незадовільно з можливістю повторного складання» / «не зараховано» FX – від 35 до 59 балів. Студент володіє матеріалом на рівні окремих фрагментів, що становлять незначну частину навчального матеріалу.

- «незадовільно з обов’язковим повторним вивченням дисципліни» / «не зараховано» F – від 1 до 34 балів. Студент не володіє навчальним матеріалом.

Таблиця відповідності результатів контролю знань за різними шкалами

100-бальною шкалою	**Шкала за ECTS**	**За національною шкалою**	
		екзамен	**залік**
90-100 (10-12)	A	Відмінно	зараховано
82-89 (8-9)	B	Добре	
74-81(6-7)	C		
64-73 (5)	D	Задовільно	
60-63 (4)	E		
35-59 (3)	Fx	незадовільно	не зараховано
1-34 (2)	F		

ПЛАН – КОНСПЕКТ ЛЕКЦІЙ ДИСЦИПЛІНИ

АНАЛІЗ ДАНИХ ТА КОМП'ЮТЕРНА СТАТИСТИКА

Комп'ютерна статистика відіграє ключову роль у сучасних дослідженнях і бізнес-аналітиці, забезпечуючи ефективні методи для аналізу та інтерпретації складних даних.

Аналіз даних та комп'ютерна статистика – це науковий напрямок, що поєднує методи обробки, аналізу, візуалізації та інтерпретації великих обсягів даних з використанням комп'ютерних засобів. Він знаходиться на перетині статистики, інформатики, математики та інших дисциплін, що займаються аналізом інформації.

ЗАСТОСУВАННЯ

Бізнес-аналітика: аналіз продажів, маркетингових кампаній, управління ризиками, моделювання поведінки споживачів.

Фінанси: прогнозування фінансових показників, оцінка ризиків, аналіз портфелів.

Медицина: аналіз ефективності лікувань, прогнозування захворювань, генетичний аналіз.

Соціальні науки: дослідження соціальних явищ, оцінка ефективності політичних рішень.

Існує широкий спектр програмних засобів для аналізу даних, таких як **R, Python (Pandas, SciPy, Statsmodels), SPSS, SAS, Stata, MATLAB**, що надають користувачам потужні інструменти для проведення статистичного аналізу, обробки великих обсягів даних, створення моделей і візуалізації результатів.

Основні етапи статистичного аналізу

1. **Планування збору даних**

Завдання: Визначення мети дослідження, вибір об'єкту та методу збору даних, проектування вибірки, розробка інструментів для збору даних (анкети, опитувальники, програми експериментів).

Методи: Суцільний та вибірковий методи збору даних.

Суцільний метод – включає в себе обробку всіх наявних даних або всієї популяції, забезпечує повну картину, але може бути дорогим і трудомістким. Використовується, коли потрібно аналізувати всю доступну інформацію.

Вибірковий метод – включає в себе обробку лише частини даних або вибіркової сукупності, що повинна бути репрезентативною для всієї популяції, економніший і швидший, ніж суцільний метод. Використовується для отримання висновків про популяцію на основі аналізу вибірки.

2. **Попередня обробка і дослідження даних**

Завдання: Очистка даних, усунення пропусків та аномалій, перевірка на відповідність нормальним розподілам, виявлення базових закономірностей у даних.

Методи: Описова статистика, побудова частотних розподілів, візуалізація даних (гістограми, коробкові діаграми, точкові діаграми).

3. **Оцінювання параметрів статистичних і математичних моделей**:

Завдання: Побудова та налаштування моделей, визначення основних параметрів, таких як середні значення, дисперсії, кореляційні коефіцієнти тощо.

Методи: Регресійний аналіз, кореляційний аналіз, дисперсійний аналіз (ANOVA), методи максимального правдоподібності, методи найменших квадратів.

4. **Перевірка сформульованих гіпотез**

Завдання: Перевірка статистичних гіпотез про зв'язки між змінними, середні значення та інші параметри, обчислення значущості результатів.

Методи: t-критерій Стьюдента, критерій Пірсона, критерій Колмогорова-Смирнова, U-критерій Манна-Уїтні, ранговий дисперсійний аналіз.

Задачі, які розв’язують методами прикладної статистики

1. **Аналіз розподілу ознак**. Визначення основних характеристик розподілу ознак (середнє значення, медіана, мода, дисперсія).
2. **Перевірка гіпотез**. Оцінка значущості відмінностей між групами або перевірка зв'язків між змінними.
3. **Регресійний аналіз**. Встановлення математичної залежності між змінними для прогнозування.
4. **Кластерний аналіз**. Групування об'єктів у класи на основі схожості між ними.

5. **Дисперсійний аналіз (ANOVA)**. Оцінка впливу одного або кількох факторів на варіативність даних.

Основні аспекти аналізу даних

1. **Збір даних**. Процес отримання даних з різних джерел, таких як опитування, експерименти, бази даних, сенсори тощо. Зібрані дані можуть бути структурованими, напівструктурованими або неструктурованими.
2. **Попередня обробка даних**.

Очищення даних. Видалення або корекція шумів, аномалій, пропущених значень, дублікатів та інших недоліків у даних.

Трансформація даних. Перетворення даних у форму, придатну для подальшого аналізу, наприклад, через нормалізацію, стандартизацію або кодування категоріальних змінних.

3. **Аналіз даних**.

Описова статистика. Перший етап аналізу даних, що включає обчислення основних показників, таких як середнє значення, медіана, мода, варіація, дисперсія тощо.

Інференційна статистика. Використовується для узагальнення результатів на основі вибірки та включає методи тестування гіпотез, побудови довірчих інтервалів та регресійний аналіз.

Машинне навчання. Використання алгоритмів для створення моделей, що можуть прогнозувати або класифікувати нові дані на основі тренувального набору даних.

4. **Візуалізація даних**. Важливий етап, який допомагає зрозуміти структуру, закономірності та аномалії в даних. Використовуються різні види графіків, діаграм, карт, що полегшує інтерпретацію результатів аналізу.
5. **Інтерпретація результатів**. Оцінка та пояснення отриманих результатів у контексті досліджуваної проблеми. Результати можуть бути використані для прийняття рішень, прогнозування або створення рекомендацій.

Алгоритми та методи:

- **Методи регресійного аналізу**. Лінійна та нелінійна регресія, логістична регресія, поліноміальна регресія тощо.

- **Кластеризація**. Алгоритми, такі як k-means, ієрархічна кластеризація, DBSCAN для сегментації даних.
- **Класифікація**. Алгоритми, включаючи дерева рішень, SVM, нейронні мережі, для побудови моделей, що розподіляють об'єкти по категоріям.
- **Тестування гіпотез**. Методи перевірки статистичних гіпотез, такі як t-критерій, U-критерій Манна-Уїтні, критерій Пірсона тощо.

Мода, медіана, середнє та варіація

1. **Мода (Mode)**

Означення: Мода — це значення, яке найчастіше зустрічається в наборі даних.

Властивості: У наборі даних може бути одна мода (унімодальний розподіл), дві моди (бімодальний розподіл) або більше (мультимодальний розподіл).

Приклад: Якщо є дані [2, 3, 3, 6, 6, 6, 7, 8], то мода = 6, оскільки це значення зустрічається найчастіше.

2. **Медіана (Median)**

Означення: Медіана — це середнє значення впорядкованого набору даних, яке ділить набір на дві рівні частини.

Обчислення:

- Якщо кількість спостережень непарна, медіана — це середнє значення в наборі даних.
- Якщо кількість спостережень парна, медіана — це середнє арифметичне двох середніх значень.

Приклад: Якщо є дані [1, 3, 3, 6, 7, 8, 9], то медіана = 6. Якщо є дані [1, 2, 3, 4, 5, 6, 8, 9], то медіана = (4+5)/2 = 4.5.

3. **Середнє (Mean)**

Означення: Середнє арифметичне (або просто середнє) — це сума всіх значень у наборі даних, поділена на кількість цих значень.

Формула: $\mu = \frac{1}{n}\sum x_i$

де x_i – кожне окреме значення, а n – кількість значень.

Приклад: Якщо є дані [2, 4, 6, 8, 10], то середнє = (2 + 4 + 6 + 8 + 10)/5 = 6.

4. **Варіація (Variance)**

Означення: Варіація — це міра розкиду значень у наборі даних відносно середнього значення. Вона показує, наскільки сильно значення відхиляються від середнього.

Формула: $V = \frac{1}{n+1}\sum(x_i - \mu)^2$

де x_i – кожне окреме значення, n – кількість значень.

Приклад: Якщо є дані [2, 4, 6, 8, 10], то середнє = 6. Варіація буде обчислена як:

$$\text{Варіація} = \frac{(2-6)^2 + (4-6)^2 + (6-6)^2 + (8-6)^2 + (10-6)^2}{4} = 10.$$

Зауваження:

1. Мода, медіана і середнє є різними способами вимірювання центральної тенденції в наборі даних.
2. Варіація показує, наскільки дані розкидані відносно цих центральних значень.
3. У симетричних розподілах (наприклад, нормальний розподіл) мода, медіана і середнє часто співпадають. У асиметричних розподілах ці значення можуть суттєво відрізнятися одне від одного.

Інтервальні оцінки

Інтервальні оцінки – це спосіб оцінювання невідомого параметра генеральної сукупності, наприклад середнього або частки, у вигляді інтервалу, в якому цей параметр знаходиться з певною ймовірністю. Інтервал, який містить невідомий параметр із заданою ймовірністю (рівнем довіри), називається довірчим інтервалом.

Основні аспекти інтервальних оцінок

Довірчий інтервал – це інтервал значень, який з певною ймовірністю (наприклад, 95%) містить справжнє значення оцінюваного параметра.

Рівень довіри (1 - α) – це ймовірність, що справжнє значення параметра міститься в довірчому інтервалі. Наприклад, рівень довіри 95% означає, що в 95%

випадків, коли ми обчислюємо довірчий інтервал за допомогою цього методу, він буде містити справжнє значення параметра.

Ширина довірчого інтервалу залежить від розміру вибірки, варіабельності даних і рівня довіри. При збільшенні вибірки або зменшенні рівня довіри ширина інтервалу зменшується.

Реалізація в статистичних пакетах

1. **SPSS**:

Після проведення аналізу, наприклад, вибіркового середнього, SPSS автоматично генерує довірчий інтервал для оцінюваного параметра.

Для отримання інтервальних оцінок середнього або частки, потрібно використовувати команду **Analyze → Descriptive Statistics → Explore** або **Analyze → Compare Means**.

2. **R**:

Функції t.test() і prop.test() дозволяють отримувати довірчі інтервали для середнього та частки відповідно.

Наприклад, для обчислення довірчого інтервалу середнього можна використовувати: t.test(x)$conf.int

Для часток: prop.test(x, n)$conf.int

3. **Python (SciPy, Statsmodels)**:

Використання бібліотеки scipy.stats або statsmodels дозволяє легко обчислити довірчі інтервали.

Наприклад: import scipy.stats as stats

```
stats.t.interval(0.95, len(data)-1, loc=np.mean(data), scale=stats.sem(data))
```

Для частки: import statsmodels.api as sm

```
sm.stats.proportion_confint(count, nobs, alpha=0.05)
```

4. **Excel**:

В Excel можна використовувати функції CONFIDENCE.T або CONFIDENCE.NORM для обчислення довірчих інтервалів.

Наприклад, для довірчого інтервалу середнього:

=CONFIDENCE.T(альфа, стандартне відхилення, розмір вибірки)

Застосування інтервальних оцінок у статистичних дослідженнях дає змогу не лише оцінити параметри сукупності, але й зрозуміти ступінь невизначеності цих оцінок. Завдяки статистичним пакетам цей процес стає автоматизованим та зручним для користувача.

Перевірка гіпотез

Критерій Стьюдента (t-критерій) — це статистичний тест, який використовується для перевірки гіпотез про середні значення однієї або двох вибірок, коли розподіл вибірки є нормальним або близьким до нормального, а розмір вибірки невеликий (звичайно, менше 30).

Типи задач, де застосовується критерій Стьюдента

1. **Перевірка середнього значення однієї вибірки (One-sample t-test)**. Застосовується, коли необхідно перевірити, чи дорівнює середнє значення вибірки заданому значенню. Наприклад, чи середній бал студентів з певного предмету дорівнює 70.
2. **Перевірка рівності середніх значень двох незалежних вибірок (Independent two-sample t-test)**. Використовується для порівняння середніх значень двох незалежних груп, наприклад, чи відрізняється середня зарплата чоловіків та жінок в одній організації.
3. **Перевірка рівності середніх значень двох залежних вибірок (Paired t-test)**. Використовується для порівняння середніх значень двох залежних або парних вибірок. Наприклад, дослідження ефективності нового препарату шляхом порівняння середніх показників здоров’я до і після лікування у однієї і тієї ж групи пацієнтів.
4. **Оцінка довірчих інтервалів для середнього значення**. t-критерій використовується для побудови довірчих інтервалів для середнього значення вибірки, коли розмір вибірки невеликий і стандартне відхилення популяції невідоме.

Приклади задач

1. **Порівняння ефективності двох методів навчання**. Є дві групи студентів, одна з яких навчається за традиційною методикою, а інша — за новою. Застосування t-критерію дозволяє визначити, чи є статистично значуща різниця в середніх балах між цими групами.
2. **Оцінка впливу дієти на вагу**. Перед початком і після закінчення дієти вимірюється вага групи людей. Застосування парного t-тесту дозволяє оцінити, чи є статистично значущі зміни у вазі після дієти.
3. **Перевірка відповідності середнього значення стандарту**. Виробник напоїв хоче перевірити, чи середній вміст цукру в новому продукті відповідає встановленому стандарту 20 грам на порцію. Для цього застосовується одновибірковий t-тест.

Критерій Стьюдента є потужним інструментом для статистичного аналізу, особливо коли дані мають нормальний розподіл і розмір вибірки невеликий. Він допомагає перевіряти гіпотези, що стосуються середніх значень, і широко використовується в різних галузях, включаючи медицину, психологію, економіку та соціальні науки.

Критерій Пірсона (χ^2-критерій) – це статистичний тест, який використовується для перевірки гіпотез про відповідність спостережуваних і теоретичних частот або для оцінки зв'язку між категоріальними змінними. Його часто застосовують у задачах, де дані представлені у вигляді таблиць спряженості (таблиць частот).

Типи задач, де застосовується критерій Пірсона

1. **Перевірка відповідності спостережуваних і теоретичних розподілів (χ^2-тест на відповідність)**. Використовується, коли необхідно перевірити, чи спостережувані частоти в певних категоріях відповідають теоретично очікуваним частотам. Наприклад, чи відповідає спостережуваний розподіл виборців певного віку теоретично очікуваному розподілу.
2. **Перевірка незалежності двох категоріальних змінних (χ^2-тест незалежності)**. Застосовується, коли потрібно перевірити, чи є залежність між двома категоріальними змінними. Наприклад, чи залежить вибір певного товару від статі покупця.
3. **Перевірка однорідності (χ^2-тест на однорідність)**. Використовується для перевірки, чи однакові розподіли певної змінної в різних вибірках. Наприклад, перевірка, чи однакові розподіли улюблених видів спорту серед студентів різних факультетів.

Приклади задач

1. **Перевірка генетичних розподілів**. В задачах генетики часто використовують критерій Пірсона для перевірки, чи спостережуваний розподіл ознак у потомстві відповідає теоретичному розподілу за законом Менделя.
2. **Аналіз маркетингових даних**. Наприклад, дослідження залежності між віковою групою та вибором певного типу продукту. Критерій Пірсона допоможе визначити, чи існує зв'язок між цими двома змінними.
3. **Перевірка рівномірності розподілу голосів на виборах**. Якщо є підозри на фальсифікації, χ^2-тест можна застосувати для перевірки відповідності спостережуваних результатів голосування теоретичному рівномірному розподілу.
4. **Оцінка ефективності рекламних кампаній**. Аналіз залежності між регіоном та ефективністю рекламної кампанії (кількість покупок після проведення кампанії в різних регіонах).

Критерій Пірсона є дуже корисним для аналізу якісних даних, зокрема для перевірки відповідності спостережуваних даних очікуваним значенням або для оцінки залежності між двома категоріальними змінними. Він широко використовується в соціальних науках, медицині, маркетингу, генетиці та інших галузях, де необхідно аналізувати взаємозв'язки між категоріальними змінними.

U-критерій Манна-Уїтні (також відомий як **тест Манна-Уїтні** або **непараметричний U-тест**), є одним із найбільш використовуваних непараметричних методів для порівняння двох незалежних вибірок. Він дозволяє визначити, чи відрізняються розподіли двох вибірок без припущення про нормальний розподіл.

Сутність U-критерію Манна-Уїтні: U-критерій Манна-Уїтні перевіряє нульову гіпотезу про те, що дві вибірки походять з однакових розподілів. Тобто, він оцінює, чи одна з вибірок має значно більші або менші значення порівняно з іншою.

Ранг змінної — це її порядковий номер після сортування всіх значень змінної в порядку зростання (або спадання). У контексті коефіцієнта кореляції Спірмена ранги використовуються для порівняння позицій відповідних значень двох змінних.

Наприклад, якщо у вас є набір значень змінної X: 5, 2, 8, 6, то після сортування в порядку зростання вони будуть мати такі ранги:

2 отримує ранг 1 (перше місце),

5 отримує ранг 2 (друге місце),

6 отримує ранг 3 (третє місце),

8 отримує ранг 4 (четверте місце).

Ранги дозволяють порівнювати розподіли двох змінних і оцінювати силу та напрямок їх монотонного зв'язку.

Приклад використання U-критерію Манна-Уїтні

Задача: Порівняти дві незалежні вибірки, що представляють результати тестування двох груп студентів (наприклад, групи, що навчаються за різними методиками).

Вибірки:

- Група 1: 85, 90, 78, 92, 88
- Група 2: 80, 82, 79, 85, 87

Кроки для проведення тесту:

1. **Об'єднання вибірок та ранжування значень:**

Об'єднуємо всі значення з обох груп.

Присвоюємо ранги всім значенням у спільній вибірці (при однакових значеннях присвоюється середній ранг).

2. **Обчислення статистики U для кожної групи:**

Визначаємо суму рангів для кожної групи.

Обчислюємо статистики U_1 та U_2за формулами:

$$U_1 = n_1 \cdot n_2 + \frac{n_1 \cdot (n_1 + 1)}{2} - R_1$$

$$U_2 = n_1 \cdot n_2 + \frac{n_2 \cdot (n_2 + 1)}{2} - R_2$$

де n_1 і n_2 – розміри вибірок, а R_1 і R_2 – суми рангів для першої та другої вибірки відповідно.

3. **Порівняння обчисленого U з критичним значенням:**

Порівнюємо мінімальне з двох значень U_1 та U_2 з критичним значенням $U_{\text{крит}}$ для відповідного рівня значущості та розмірів вибірок (значення беруться з таблиць критичних значень U-критерію Манна-Уїтні).

4. **Інтерпретація результатів:**

Якщо обчислене U менше або дорівнює критичному значенню, нульова гіпотеза відхиляється, що свідчить про статистично значущу різницю між вибірками.

Важливі моменти:

U-критерій Манна-Уїтні є непараметричним тестом, тому його можна використовувати навіть тоді, коли дані не відповідають нормальному розподілу.

Тест особливо корисний у випадках, коли вибірки невеликі або є підозра на відхилення від нормальності.

Приклад обчислення

Для наведених вище даних:

Об'єднана вибірка: 78, 79, 80, 82, 85, 85, 87, 88, 90, 92

Ранги: 1, 2, 3, 4, 5.5, 5.5, 7, 8, 9, 10

Сума рангів для групи 1: R_1=5.5+8+9+10+7=39.5

Сума рангів для групи 2: R_2=1+2+3+4+5.5=15.5

Обчислимо U: U_1=5·5+5·62−39.5=12.5

U_2=5·5+5·62−15.5=36.5

Мінімальне U = 12.5

Порівнюємо з критичним значенням (наприклад, для α=0.05) і робимо висновок.

Для того щоб зробити остаточний висновок, потрібно порівняти обчислене значення U з критичним значенням U із таблиці критичних значень для U-критерію Манна-Уїтні.

Допустимо, що рівень значущості α=0.05 і розміри вибірок n_1 =5 та n_2 =5. Для цього випадку критичне значення U із таблиці становить 2.

Висновок:

Якщо обчислене *U* (у нашому випадку 12.5) більше критичного значення (2), то ми не відхиляємо нульову гіпотезу. Це означає, що немає статистично значущої різниці між двома вибірками.

Якщо б обчислене *U* було меншим або дорівнювало критичному значенню (наприклад, 2 або менше), ми б відхилили нульову гіпотезу і сказали б, що між вибірками є статистично значуща різниця.

В даному випадку обчислене значення *U* (12.5) більше критичного значення (2), тому **немає підстав стверджувати, що між двома групами студентів є статистично значуща різниця у їхніх результатах**.

Дисперсійний аналіз і його реалізація в статистичних пакетах

Дисперсійний аналіз є потужним інструментом для оцінки значущості відмінностей між групами. Завдяки широкому спектру статистичних пакетів, таких як R, Python, SPSS, SAS, MATLAB, дослідники мають змогу ефективно реалізувати різні види дисперсійного аналізу залежно від специфіки своїх даних і поставлених завдань.

Дисперсійний аналіз (ANOVA, Analysis of Variance) — це статистичний метод, який використовується для порівняння середніх значень кількох груп і визначення, чи є статистично значущі відмінності між ними. Він часто застосовується в експериментальних дослідженнях для аналізу впливу одного або декількох факторів на залежну змінну.

Основні типи дисперсійного аналізу

1. **Однофакторний Дисперсійний Аналіз (One-way ANOVA)**

Призначення: Використовується для порівняння середніх значень більше ніж двох груп на основі одного фактора.

Приклад: Аналіз середньої продуктивності студентів у різних групах, що вивчають різні методи навчання.

2. **Двофакторний Дисперсійний Аналіз (Two-way ANOVA)**

Призначення: Використовується для порівняння середніх значень між групами з урахуванням двох факторів. Може враховувати взаємодію між факторами.

Приклад: Аналіз середньої продуктивності студентів з урахуванням методу навчання і типу завдання.

3. **ANOVA з повторними вимірами (Repeated Measures ANOVA)**

Призначення: Використовується, коли одні й ті самі об'єкти дослідження вимірюються кілька разів.

Приклад: Аналіз зміни ваги пацієнтів під час лікування в різні моменти часу.

4. **Множинний Дисперсійний Аналіз (MANOVA, Multivariate ANOVA)**

Призначення: Розширення ANOVA для випадків, коли є більше однієї залежної змінної.

Приклад: Аналіз впливу різних методів навчання на продуктивність і мотивацію студентів одночасно.

Реалізація дисперсійного аналізу в статистичних пакетах

1. **R**

Основні функції:

aov() — для виконання однофакторного та двофакторного дисперсійного аналізу.

anova() — для порівняння моделей або виконання двофакторного аналізу.

lme() з пакету nlme — для аналізу з повторними вимірами.

Приклад використання:

```
data <- data.frame(
  group = factor(rep(1:3, each = 10)),
  score = c(23, 25, 27, 24, 22, 26, 23, 21, 24, 26,
            30, 29, 32, 31, 30, 28, 29, 30, 28, 31,
            35, 34, 33, 37, 36, 34, 35, 32, 36, 38)
)
model <- aov(score ~ group, data = data)
summary(model)
```

2. **Python (scipy, statsmodels)**

Основні модулі:

scipy.stats.f_oneway — для виконання однофакторного дисперсійного аналізу.

statsmodels.formula.api.ols і statsmodels.stats.anova.anova_lm — для розширеного аналізу.

Приклад використання:

```
import pandas as pd
import statsmodels.api as sm
from statsmodels.formula.api import ols

data = pd.DataFrame({
   'group': ['A']*10 + ['B']*10 + ['C']*10,
   'score': [23, 25, 27, 24, 22, 26, 23, 21, 24, 26,
             30, 29, 32, 31, 30, 28, 29, 30, 28, 31,
             35, 34, 33, 37, 36, 34, 35, 32, 36, 38]
})

model = ols('score ~ C(group)', data=data).fit()
anova_table = sm.stats.anova_lm(model, typ=2)
print(anova_table)
```

3. **SPSS**

Основні функції:

Analyze -> Compare Means -> One-Way ANOVA — для однофакторного дисперсійного аналізу.

Analyze -> General Linear Model -> Univariate — для двофакторного аналізу.

Analyze -> General Linear Model -> Repeated Measures — для аналізу з повторними вимірами.

4. **SAS**

PROC ANOVA — для однофакторного дисперсійного аналізу.

PROC GLM — для виконання складних моделей дисперсійного аналізу, включаючи повторні виміри.

Приклад використання:

```
data scores;

  input group $ score;
  datalines;
A 23
A 25
A 27
A 24
A 22
B 30
B 29
B 32
B 31
B 30
C 35
C 34
C 33
C 37
C 36
;
run;

proc anova data=scores;
  class group;
  model score = group;
  means group;
run;
```

5. **MATLAB**

Основні функції:

anova1() — для однофакторного аналізу.

anova2() — для двофакторного аналізу.

fitrm() — для аналізу з повторними вимірами.

Приклад використання:

```
group = [ones(10,1); 2*ones(10,1); 3*ones(10,1)];
score = [23, 25, 27, 24, 22, 26, 23, 21, 24, 26, ...
        30, 29, 32, 31, 30, 28, 29, 30, 28, 31, ...
        35, 34, 33, 37, 36, 34, 35, 32, 36, 38];
p = anova1(score, group);
```

Кореляційний аналіз і його реалізація в статистичних пакетах

Кореляційний аналіз є статистичним методом, який досліджує взаємозв'язок між двома або більше змінними. Основною метою цього аналізу є виявлення та кількісна оцінка сили і напряму зв'язку між змінними.

Основні методи кореляційного аналізу

1. **Парний Кореляційний Аналіз**

Коефіцієнт кореляції Пірсона. Використовується для оцінки лінійного зв'язку між двома кількісними змінними. Величина коефіцієнта кореляції Пірсона (r) варіюється від -1 до 1.

Значення r = 1 означає ідеальну позитивну лінійну залежність.

Значення r = -1 означає ідеальну негативну лінійну залежність.

Значення r = 0 свідчить про відсутність лінійного зв'язку.

Коефіцієнт кореляції Пірсона визначається за такою формулою:

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2}}$$

Де: r – коефіцієнт кореляції Пірсона;

x_i і y_i – значення змінних *x* та *y* для *i*-го спостереження;

$\bar{x}$ і $\bar{y}$ – середні значення змінних *x* та *y*;

n – кількість спостережень.

Ця формула оцінює лінійний зв'язок між двома змінними. Значення коефіцієнта кореляції Пірсона може бути від -1 до 1, де:

r=1 вказує на ідеальний позитивний лінійний зв'язок,

r=−1 вказує на ідеальний негативний лінійний зв'язок,

r=0 вказує на відсутність лінійного зв'язку.

Коефіцієнт кореляції Спірмена. Використовується для оцінки монотонного зв'язку між двома змінними, незалежно від того, чи є зв'язок лінійним. Часто використовується, коли дані не відповідають нормальному розподілу.

Коефіцієнт кореляції Спірмена обчислюється за такою формулою:

$$r_s = 1 - \frac{6 \sum_{i=1}^{n} d_i^2}{n(n^2 - 1)}$$

Де r_s – коефіцієнт кореляції Спірмена;

d_i – різниця між рангами відповідних значень x_i і y_i;

n — кількість спостережень.

Коефіцієнт кореляції Спірмена використовується для оцінки монотонного зв'язку між двома змінними. Його значення також може бути від -1 до 1, де

r_s=1 вказує на ідеальний позитивний монотонний зв'язок,

r_s=−1 вказує на ідеальний негативний монотонний зв'язок,

r_s=0 вказує на відсутність монотонного зв'язку.

Коефіцієнт кореляції Кендалла. Використовується для вимірювання зв'язку між ранговими змінними. Часто застосовується при невеликій кількості спостережень або при наявності значних викидів у даних.

Коефіцієнт кореляції Кендалла τ використовується для оцінки ступеня асоціації між двома ранжованими змінними. Він визначається як різниця між імовірністю того, що спостереження мають однаковий порядок (конкордантні пари) та імовірністю того, що вони мають протилежний порядок (дискордантні пари).

Формула коефіцієнта кореляції Кендалла виглядає наступним чином:

$$\tau = \frac{C - D}{\frac{1}{2}n(n-1)}$$

Де C – кількість конкордантних пар (пар, які мають однаковий порядок);

D – кількість дискордантних пар (пар, які мають протилежний порядок);

n – кількість спостережень.

Значення коефіцієнта Кендалла варіюється від -1 до 1, де:

τ=1 означає повну позитивну кореляцію,

τ=−1 означає повну негативну кореляцію,

τ=0 означає відсутність кореляції.

Коефіцієнт кореляції Кендалла τ використовується для оцінки ступеня асоціації між двома ранжованими змінними. Він визначається як різниця між імовірністю того, що спостереження мають однаковий порядок (конкордантні пари) та імовірністю того, що вони мають протилежний порядок (дискордантні пари).

Множинний кореляційний аналіз визначає зв'язок між кількома незалежними змінними та однією залежною змінною.

Кореляційний аналіз є важливим інструментом для вивчення взаємозв'язку між змінними в різних галузях досліджень. Завдяки його широкому застосуванню в різних статистичних пакетах, таких як R, Python, SPSS, SAS, MATLAB, дослідники можуть легко реалізовувати і аналізувати результати кореляційного аналізу залежно від своїх завдань та специфіки даних.

Реалізація кореляційного аналізу в статистичних пакетах

1. **R**

Основні функції:

cor() — для обчислення коефіцієнта кореляції Пірсона, Спірмена та Кендалла.

cor.test() — для виконання тестів на значущість кореляції.

Приклад використання:

```
data <- data.frame(
  x = c(10, 20, 30, 40, 50),
  y = c(15, 25, 35, 45, 55)
)
cor(data$x, data$y, method = "pearson")
cor.test(data$x, data$y, method = "pearson")
```

2. **Python (scipy, pandas, statsmodels)**

Основні модулі:

scipy.stats.pearsonr — для обчислення коефіцієнта кореляції Пірсона.

scipy.stats.spearmanr — для обчислення коефіцієнта кореляції Спірмена.

scipy.stats.kendalltau — для обчислення коефіцієнта кореляції Кендалла.

Приклад використання:

```
import pandas as pd
from scipy.stats import pearsonr, spearmanr, kendalltau

data = pd.DataFrame({
    'x': [10, 20, 30, 40, 50],
    'y': [15, 25, 35, 45, 55]
})

pearson_corr, _ = pearsonr(data['x'], data['y'])
spearman_corr, _ = spearmanr(data['x'], data['y'])
kendall_corr, _ = kendalltau(data['x'], data['y'])
```

```
print(f"Pearson correlation: {pearson_corr}")
print(f"Spearman correlation: {spearman_corr}")
print(f"Kendall correlation: {kendall_corr}")
```

3. **SPSS**

Основні функції:

Analyze -> Correlate -> Bivariate — для виконання парного кореляційного аналізу.

Analyze -> Correlate -> Partial — для виконання часткової кореляції.

Analyze -> Correlate -> Multiple — для виконання множинної кореляції.

Приклад використання: В SPSS, після вибору опції Bivariate, користувачеві потрібно вибрати змінні, метод кореляції (наприклад, Пірсон або Спірмен) та інші налаштування, після чого програма автоматично виконає аналіз.

4. **SAS**

PROC CORR — основна процедура для обчислення коефіцієнтів кореляції.

Приклад використання:

```
data example;
  input x y;
  datalines;
10 15
20 25
30 35
40 45
50 55
;
run;

proc corr data=example;
  var x y;
run;
```

5. **MATLAB**

Основні функції:

corr() — для обчислення коефіцієнтів кореляції Пірсона, Спірмена та Кендалла.

Приклад використання:

```
x = [10 20 30 40 50];
y = [15 25 35 45 55];
R = corr(x', y', 'Type', 'Pearson')
```

Регресійний аналіз

Регресійний аналіз – це метод статистичного аналізу, який використовується для дослідження залежності однієї або кількох залежних змінних (вихідних, респондентів) від однієї або кількох незалежних змінних (предикторів). Основна мета регресійного аналізу — оцінити взаємозв'язки між змінними і створити модель, яка дозволяє прогнозувати значення залежної змінної на основі значень незалежних змінних.

Основні етапи регресійного аналізу:

1. **Вибір моделі**. Визначення форми регресійного рівняння (лінійна, поліноміальна, логістична тощо).
2. **Оцінка параметрів.** Використання методу найменших квадратів (OLS) для оцінки коефіцієнтів регресійного рівняння.
3. **Аналіз значущості моделі**. Перевірка статистичної значущості регресійних коефіцієнтів та моделі в цілому (наприклад, за допомогою F-тесту, t-тесту).
4. **Перевірка припущень**. Аналіз залишків, перевірка на гетероскедастичність, мультиколінеарність, автокореляцію та нормальність залишків.
5. **Прогнозування**. Використання побудованої моделі для прогнозування значень залежної змінної.

Реалізація регресійного аналізу в статистичних пакетах

У різних статистичних пакетах регресійний аналіз реалізований за допомогою спеціалізованих функцій та команд:

1. **R**:

Для лінійної регресії використовується функція lm():

```
model <- lm(y ~ x1 + x2, data = dataset)
summary(model)
```

Для поліноміальної регресії можна використовувати ту ж функцію, але з поліноміальним розширенням:

```
model <- lm(y ~ poly(x1, 2) + x2, data = dataset)
```

2. **Python (пакет statsmodels або scikit-learn)**:

У statsmodels:

```
import statsmodels.api as sm
model = sm.OLS(y, X).fit()
model.summary()
```

У scikit-learn:

```
from sklearn.linear_model import LinearRegression
model = LinearRegression().fit(X, y)
model.coef_
```

3. **SPSS**:

Регресійний аналіз доступний через меню: **Analyze → Regression → Linear...**

Можна вибрати залежні та незалежні змінні, а також переглянути результати оцінки моделі.

4. **Stata**:

Використовується команда regress: regress y x1 x2

Приклад використання регресійного аналізу

Задача: Вивчити залежність витрат на рекламу та продажів у компанії.

1. Зібрані дані: залежна змінна Sales (продажі) та незалежна змінна AdSpend (витрати на рекламу).
2. Вибір моделі: лінійна регресія, де Sales залежить від AdSpend.
3. Оцінка параметрів:

```
model <- lm(Sales ~ AdSpend, data = dataset)
```

summary(model)

4. Результати:

Отримано коефіцієнт для AdSpend, який показує, як змінюються продажі при збільшенні витрат на рекламу на одиницю.

Статистична значущість моделі оцінена за допомогою p-значень.

Регресійний аналіз є одним з ключових інструментів статистичного аналізу даних і широко використовується у різних галузях, включаючи економіку, фінанси, соціологію та маркетинг.

Логістична регресія — це метод статистичного аналізу, який використовується для прогнозування результату змінної, що має двійковий або категорійний характер (залежна змінна), на основі однієї або декількох незалежних змінних. Це один з найпоширеніших методів у аналізі даних, особливо в ситуаціях, де результатами можуть бути такі значення, як "так/ні", "успіх/невдача", "0/1".

Основні аспекти логістичної регресії:

1. **Модель логістичної регресії**. В логістичній регресії використовують логістичну функцію для моделювання ймовірності, що залежна змінна прийме певне значення:

$$P(Y = 1) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 X_1 + \cdots + \beta_n X_n)}}$$

Тут β_0 – це вільний член, а $\beta_0, \ldots, \beta_n$ – коефіцієнти незалежних змінних.

2. **Оцінка параметрів**. Коефіцієнти моделі оцінюються методом максимальної правдоподібності.
3. **Інтерпретація коефіцієнтів**. Коефіцієнти в логістичній регресії можна інтерпретувати як вплив незалежних змінних на логарифм шансу настання певного результату.
4. **Тестування моделі**. Для перевірки моделі використовуються такі статистичні критерії, як тест Вальда, критерій хі-квадрат (χ^2), коефіцієнт детермінації псевдо- R^2 та інші.

Реалізація логістичної регресії в статистичних пакетах

1. **R**:

Функція glm() (генералізована лінійна модель) з параметром family=binomial використовується для виконання логістичної регресії.

Приклад:

```
model <- glm(Y ~ X1 + X2 + X3, family = binomial, data = dataset)
summary(model)
```

Бібліотеки, такі як caret, ROCR, pROC, можуть бути використані для оцінки якості моделі та побудови ROC-кривих.

2. **Python**:

Бібліотека statsmodels (функція Logit() для бінарної логістичної регресії).

Бібліотека sklearn (функція LogisticRegression для простої та багаторазової логістичної регресії).

Приклад:

```
from sklearn.linear_model import LogisticRegression
model = LogisticRegression()
model.fit(X, y)
```

Інструменти, такі як metrics з sklearn, використовуються для оцінки моделі.

3. **SPSS**:

Меню: Analyze > Regression > Binary Logistic.

SPSS надає зручний інтерфейс для вибору незалежних змінних, а також для інтерпретації результатів, включаючи оцінку коефіцієнтів та побудову таблиць контингентності.

4. **SAS**:

Процедура PROC LOGISTIC використовується для проведення логістичної регресії.

SAS також дозволяє проводити стратифікацію за групами, аналіз залишків та інші розширені аналізи.

5. **Stata**:

Команда logit для виконання бінарної логістичної регресії.

Команда logistic для виводу ймовірностей.

Stata надає широкий набір інструментів для аналізу результатів, побудови прогнозів та інтерпретації моделі.

6. **MATLAB**:

Функція mnrfit для багатономінальної логістичної регресії.

MATLAB також підтримує функції для візуалізації результатів та оцінки моделі.

Приклади використання логістичної регресії

Медицина. Прогнозування наявності захворювання на основі таких чинників, як вік, індекс маси тіла, рівень холестерину тощо.

Соціальні науки. Аналіз ймовірності голосування за певного кандидата в залежності від віку, освіти, доходу тощо.

Маркетинг. Прогнозування ймовірності купівлі продукту на основі поведінкових характеристик клієнтів.

Логістична регресія є потужним інструментом для аналізу ймовірностей у різних галузях, і її реалізація у статистичних пакетах дозволяє легко застосовувати цей метод до реальних даних.

Багатовимірний статистичний аналіз

Кластерний аналіз – це метод багатовимірного статистичного аналізу, який використовується для групування об'єктів або спостережень у класи, або кластери, на основі їхніх властивостей або ознак. Основна мета кластерного аналізу — знайти природні групи (кластери) в даних таким чином, щоб об'єкти в одному кластері були більш схожими один на одного, ніж на об'єкти з інших кластерів.

Основні методи кластерного аналізу

1. **Метод k-середніх (k-means)**. Один з найбільш поширених алгоритмів кластеризації, який працює шляхом поділу даних на k кластерів. Кожен кластер представляється своїм центром (середнім значенням), і об'єкти приписуються до найближчого центру. Алгоритм ітеративно оновлює центри кластерів до тих пір, поки центри не стабілізуються.

2. **Ієрархічна кластеризація**. Створює дерево кластерів (дендрограму), де всі об'єкти починають в окремих кластерах, і поступово об'єднуються в більші кластери. Може бути агломеративною (злиття малих кластерів у більші) або дивізивною (поділ великих кластерів на менші).
3. **Метод середньої щільності (DBSCAN)**. Виявляє кластери будь-якої форми на основі щільності об'єктів у просторі. Він також ідентифікує шуми (дані, які не належать до жодного кластера). Особливо корисний для виявлення аномалій або виявлення кластерів з нерегулярною формою.
4. **Метод головних компонент (PCA) з подальшою кластеризацією**. Спочатку знижує розмірність даних, виділяючи основні компоненти, що пояснюють найбільшу варіативність, а потім застосовує кластеризацію до зменшених даних.

Реалізація кластерного аналізу в статистичних пакетах:

1. **R**

Бібліотеки: stats, cluster, factoextra, NbClust.

Приклади функцій: kmeans(), hclust(), dbscan().

Візуалізація: fviz_cluster() для візуалізації результатів кластеризації.

2. **Python**

Бібліотеки: scikit-learn, SciPy, Matplotlib.

Приклади функцій: KMeans, AgglomerativeClustering, DBSCAN.

Візуалізація: matplotlib, seaborn, plotly для візуалізації кластерів і дендрограм.

3. **SPSS**

Інструмент Hierarchical Cluster Analysis для ієрархічної кластеризації.

Інструмент K-Means Cluster Analysis для застосування методу k-середніх.

Інструмент TwoStep Cluster Analysis для автоматичної кластеризації великих наборів даних.

4. **SAS**

Процедура PROC CLUSTER для ієрархічної кластеризації.

Процедура PROC FASTCLUS для кластеризації методом k-середніх.

5. **Stata**

Команди: cluster kmeans, cluster hierarchical, cluster stop.

Інструменти для аналізу результатів кластеризації та їх візуалізації.

6. **MATLAB**

Функції: kmeans(), linkage() для ієрархічної кластеризації, clusterdata() для автоматичної кластеризації.

Використання кластерного аналізу

- **Сегментація ринку**. Поділ споживачів на групи на основі їхньої поведінки, уподобань, соціально-демографічних характеристик.
- **Аналіз геномних даних**. Групування генів або експресій генів на основі схожості в їхніх профілях.
- **Виявлення аномалій**. Виявлення відхилень у даних, які можуть бути ознаками шахрайства, дефектів або збоїв у системах.
- **Профілювання клієнтів**. Групування клієнтів банку або страхової компанії на основі їхніх фінансових дій, ризиків або платіжної історії.

Кластерний аналіз є важливим інструментом для виявлення структур у великих наборах даних і знаходження схожих об'єктів, що дозволяє приймати більш обґрунтовані рішення у різних галузях, від маркетингу до біоінформатики.

Дискримінантний аналіз

Дискримінантний аналіз – це метод статистичного аналізу, який використовується для розділення та класифікації об'єктів на різні групи (класи) на основі їхніх характеристик. Основна мета дискримінантного аналізу – знайти лінійні комбінації змінних, які найкраще розділяють групи. Ці лінійні комбінації називаються дискримінантними функціями.

Основні етапи дискримінантного аналізу

1. **Визначення змінних і груп**. Вибираються змінні (ознаки), які будуть використовуватись для розділення об'єктів на групи. Визначаються групи (класи), між якими необхідно провести дискримінацію.

2. **Побудова дискримінантних функцій**. Створюються лінійні комбінації змінних, які максимізують різницю між групами. Для двох груп використовується одна дискримінантна функція, для більшої кількості груп може бути побудовано кілька функцій.
3. **Оцінка якості класифікації**.Оцінюється, наскільки ефективно дискримінантні функції розділяють групи. Використовуються різні критерії, такі як процент правильно класифікованих об'єктів або канонічні кореляції.
4. **Перевірка на нових даних**. Модель перевіряється на нових даних для оцінки її здатності до узагальнення.

Реалізація дискримінантного аналізу в статистичних пакетах

1. **R**

Бібліотеки: MASS (функція lda() для лінійного дискримінантного аналізу), MASS (функція qda() для квадратичного дискримінантного аналізу).

Інші бібліотеки: candisc для канонічного дискримінантного аналізу, klaR для розширених методів дискримінантного аналізу.

2. **Python**

Бібліотеки: scikit-learn.

Функції: LinearDiscriminantAnalysis для лінійного дискримінантного аналізу, QuadraticDiscriminantAnalysis для квадратичного дискримінантного аналізу.

Візуалізація: matplotlib, seaborn для візуалізації результатів аналізу.

3. **SPSS**

Меню: Analyze > Classify > Discriminant.

SPSS дозволяє обирати змінні для аналізу, створювати дискримінантні функції та класифікувати нові об'єкти на основі побудованої моделі.

4. **SAS**

Процедура PROC DISCRIM.

Використовується для виконання лінійного і квадратичного дискримінантного аналізу та побудови функцій, що розділяють групи.

5. **Stata**

Команда discrim lda для лінійного дискримінантного аналізу.

Команда discrim qda для квадратичного дискримінантного аналізу.

6. **MATLAB**

Функції: fitcdiscr для побудови дискримінантних моделей.

Дозволяє виконувати як лінійний, так і квадратичний дискримінантний аналіз.

Приклади використання дискримінантного аналізу

- **Класифікація споживачів**. Розподіл клієнтів на групи на основі їхньої поведінки або соціально-демографічних характеристик.
- **Медична діагностика**. Виявлення наявності певної хвороби на основі медичних показників.
- **Фінансовий аналіз**. Розподіл компаній на групи (наприклад, успішні/неуспішні) на основі фінансових показників.
- **Аналіз ризиків**. Оцінка ризику дефолту на основі фінансових характеристик боржників.

Дискримінантний аналіз є потужним інструментом для задач класифікації, особливо коли є достатньо даних для побудови надійних моделей і коли групи мають значні відмінності між собою.

Факторний аналіз

Факторний аналіз – це метод багатовимірного статистичного аналізу, який використовується для виявлення прихованих факторів або змінних, що пояснюють структуру даних. Основна мета факторного аналізу полягає в зменшенні кількості змінних і визначенні підгруп змінних, які є взаємопов'язаними, тобто мають спільний фактор.

Основні етапи факторного аналізу

1. **Вибір змінних для аналізу**. Визначаються вихідні змінні, які будуть аналізуватися. Вони повинні бути числовими і відповідати певним умовам (наприклад, нормальність розподілу).
2. **Побудова кореляційної матриці**. Обчислюється кореляційна матриця змінних, щоб оцінити ступінь взаємозв'язку між ними.

3. **Вибір методу факторизації**. Найпопулярнішими методами є метод головних компонент (Principal Component Analysis, PCA) і метод максимальної правдоподібності (Maximum Likelihood, ML).
4. **Обертання факторів**. Після визначення факторів часто проводиться обертання факторів (варімакс, промакс) для спрощення інтерпретації факторів.
5. **Інтерпретація факторів**. Визначення, які змінні найбільше впливають на кожен фактор, і інтерпретація факторів на основі цих змінних.
6. **Перевірка адекватності моделі**. Оцінка моделі на основі різних статистичних критеріїв (наприклад, критерій Кайзера-Мейєра-Олкіна (KMO), тест Бартлетта).

Реалізація факторного аналізу в статистичних пакетах

1. **R**

Бібліотека psych (функція fa() для факторного аналізу).

Бібліотека factoextra для візуалізації результатів факторного аналізу.

Функція principal() з бібліотеки psych для головних компонент.

2. **Python**

Бібліотека sklearn.decomposition (функція FactorAnalysis для факторного аналізу).

Бібліотека FactorAnalyzer для розширеного аналізу, включаючи обертання факторів.

3. **SPSS**

Меню: Analyze > Dimension Reduction > Factor.

SPSS дозволяє обирати кількість факторів, метод факторизації та обертання, а також надає інструменти для інтерпретації факторів.

4. **SAS**

Процедура PROC FACTOR.

Використовується для виконання різних видів факторного аналізу, включаючи метод головних компонент та обертання факторів.

5. **Stata**

Команда factor для виконання факторного аналізу.

Команда rotate для обертання факторів.

6. **MATLAB**

Функція factoran для факторного аналізу.

MATLAB підтримує різні методи факторного аналізу та надає можливості для візуалізації результатів.

Приклади використання факторного аналізу

- **Соціологічні дослідження**. Виявлення основних факторів, що впливають на суспільні думки чи поведінку людей на основі великої кількості опитувальних змінних.
- **Психометрія**. Виявлення латентних факторів, що впливають на результати тестів, таких як інтелектуальні чи особистісні тести.
- **Економіка**. Визначення макроекономічних факторів, які впливають на показники, такі як інфляція, безробіття або ВВП.
- **Маркетингові дослідження**. Виявлення факторів, які впливають на купівельні рішення, на основі споживчих переваг.

Факторний аналіз є корисним інструментом для аналізу складних наборів даних і виявлення прихованих структур, що допомагає спростити моделі та зробити їх більш інтерпретованими.

Спільні риси та відмінності між кластерним, дискримінантним та факторним аналізами

Спільні риси:

1. **Призначення для багатовимірних даних**. Усі три методи аналізу призначені для роботи з багатовимірними даними, де існує більше однієї змінної, які взаємодіють між собою.
2. **Зниження розмірності**. Як кластерний, так і факторний аналіз можуть використовуватися для зниження розмірності даних, спрощення їхньої структури та виділення основних тенденцій.
3. **Інструменти для аналізу прихованих структур**. Факторний та кластерний аналізи виявляють приховані структури в даних: кластерний аналіз групує дані за схожістю, а факторний аналіз виявляє приховані фактори, що впливають на змінні.

4. **Використання у практичних задачах**. Всі три методи активно застосовуються в різних галузях, таких як маркетинг, соціологія, психологія та економіка, для вирішення конкретних практичних задач.
5. **Візуалізація результатів**. Усі три методи передбачають можливість візуалізації результатів (дендрограми для кластерного аналізу, дискримінантні функції для дискримінантного аналізу, факторні навантаження для факторного аналізу).

Відмінності:

1. **Ціль аналізу**.

Кластерний аналіз використовується для поділу набору даних на групи (кластери) на основі схожості спостережень.

Дискримінантний аналіз використовується для класифікації спостережень у заздалегідь визначені групи (класи) та для виявлення найкращих предикторів для цієї класифікації.

Факторний аналіз використовується для виявлення прихованих факторів, які пояснюють кореляції між спостережуваними змінними.

2. **Тип аналізу**.

Кластерний аналіз є методом некерованого навчання, де групи (кластери) утворюються на основі схожості даних без попередньо відомих міток.

Дискримінантний аналіз є методом керованого навчання, де існують попередньо визначені класи, і завдання полягає в тому, щоб знайти функції, які найкраще розділяють ці класи.

Факторний аналіз не є методом класифікації чи групування, а призначений для зменшення кількості змінних та виявлення спільних факторів.

3. **Методологія**.

Кластерний аналіз базується на метриках схожості або відстані між об'єктами (наприклад, Євклідова відстань).

Дискримінантний аналіз базується на побудові лінійних (або нелінійних) функцій, що розділяють простір спостережень на класи.

Факторний аналіз базується на кореляційних матрицях і спрямований на пояснення варіативності даних за допомогою меншої кількості факторів.

4. **Вхідні дані**.

Кластерний аналіз потребує тільки матрицю ознак і не вимагає наявності попередньо визначених груп.

Дискримінантний аналіз вимагає наявності матриці ознак та вектору міток класів.

Факторний аналіз працює лише з матрицею ознак і не передбачає жодних груп чи класів.

5. **Вихідні результати**.

Кластерний аналіз: результатом є поділ даних на кілька кластерів, кожен з яких має свою характеристику.

Дискримінантний аналіз: результатом є дискримінантні функції, які максимально відрізняють класи.

Факторний аналіз: результатом є фактори та їхні навантаження на початкові змінні.

Ці спільні риси та відмінності допомагають зрозуміти, як кожен метод підходить до аналізу багатовимірних даних і яку роль він відіграє у дослідженнях.

РЕКОМЕНДОВАНА ЛІТЕРАТУРА

Основна

1. Стрелковська І.В. Математична статистика /І.В. Стрелковська, В.М. Паскаленко. – Одеса: ОНАЗ, 2019. – 110с.

2. Майборода Р.Є. Комп'ютерна статистика. Професійний старт. Навчальний посібник. «Київський університет», 2018. – 482 с.

3. Майборода Р.Є., Сугакова О.В. Аналіз даних за допомогою R. – Навчальний посібник. «Київський університет», 2015. – 65 с.

4. Стрелковська І.В., Григор'єва Т.І. Математичні методи в наукових дослідженнях: методичні рекомендації для самостійної роботи здобувачів / 2023, МГУ, Одеса, 22с.

Допоміжна

5. Стрелковська І.В. Теорія ймовірностей та випадкові процеси (для фахівців у галузі ІТ-технологій) / І.В. Стрелковська, В.М. Паскаленко. – Одеса: ОНАЗ, 2018. – 384 с.

6. Оленко А.Я. Комп'ютерна статистика. Навчальний посібник.— К., ВПЦ "Київський університет", 2007. -174с.

7. Стрелковська І.В. Вища математика для фахівців в галузі зв'язку. Ч.5 / І.В. Стрелковська, В.М. Паскаленко. – Одеса: ВМВ, 2018. – 508 с.

8. *Thomas Rahlf: Data Visualisation with R*: Springer International Publishing, 2017, 385 p.

9. Гнатюк В. Вступ до R на прикладах: навчальний посібник.- Навчальний посібник. ХНЕУ, 2010, 107с.

10. Стрелковська І.В., Золотухін Р.В., Григор'єва Т.І. Узагальнена модель оцінки показників функціонування низькошвидкісних мереж зв'язку автоматизованих систем управління. Інфокомунікаційні та комп'ютерні технології. – 2022. – № 1 (03). – С. 138-153**.**

11. Мішура Ю. С. Випадкові процеси: теорія, статистика, застосування : підручник / Ю. С. Мішура, К. В. Ральченко, Г. М. Шевченко. – 2-ге вид., випр. і допов. – К. : ВПЦ "Київський університет", 2021. – 496 с.

12. Випадкові процеси: теорія, статистика, застосування : підручник / Ю. С. Мішура, К. В. Ральченко, Л. М. Сахно, Г. М. Шевченко. – 2-ге вид., випр. і допов. – К. : ВПЦ "Київський університет", 2023. – 496 с.

13. Методика та організація наукових досліджень : Навч. посіб. / С. Е. Важинський, Т І. Щербак. – Суми: СумДПУ імені А. С. Макаренка, 2016. – 260 с.

14. Strelkovskaya I., Solovskaya I., Makoganiuk A. Different extrapolation methods in Problems of Forecasting. Advances in Information and Communication Technology and Systems. Lecture Notes in Networks and Systems. 2020. Vol. 152. Springer. P. 217-228. https://doi.org/10.1007/ 978-3-030-58359-0_12

15. Strelkovskaya I., Solovskaya I., Strelkovska J. Fingerprinting/Indoor positioning using complex planar splines. Journal of Electrical Engineering. Vol. 72 (2021), N06, pp. 401-406. https://doi.org/10.2478/jee-2021-0057

16. Маценко В.Г. Математичне моделювання: навчальний посібник – Чернівці: Чернівецький національний університет, 2014.–519 с.

Інформаційні ресурси

17. Національна бібліотека України імені В. І. Вернадського [Електронний ресурс]: [Веб-сайт]. – Електронні дані. – Київ: НБУВ, 2013-2015. – Режим доступу: www.nbuv.gov.ua

18. Електронний каталог Національної парламентської бібліотеки України [Електронний ресурс]: [політемат. база даних містить відом. про вітчизн. та зарубіж. кн., брош., що надходять у фонд НПБ України]. – Електронні дані (803 438 записів). – Київ: Нац. парлам. б-ка України, 2002-2015. – Режим доступу: catalogue.nplu.org .

19. Український інститут інтелектуальної власності [Електронний ресурс]: [Веб-сайт]. – Електронні дані. – Київ: УІІВ, 2017. – Режим доступу: http://www.uipv.org

ЗМІСТ

Навчальне видання

Григор’єва Тетяна Ігорівна
Соловська Ірина Миколаївна

АНАЛІЗ ДАНИХ ТА КОМП’ЮТЕРНА СТАТИСТИКА

Методичні рекомендації для самостійної роботи здобувачів

Українською мовою