

Пасічник Н. О.
Ріжняк Р. Я.

МАТЕМАТИКО-СТАТИСТИЧНІ МЕТОДИ В ІСТОРИЧНИХ ДОСЛІДЖЕННЯХ

НАВЧАЛЬНИЙ ПОСІБНИК

Кропивницький
2025

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
Центральноукраїнський державний університет
імені Володимира Винниченка

Пасічник Н.О.
Ріжняк Р.Я.

**МАТЕМАТИКО-СТАТИСТИЧНІ МЕТОДИ
В ІСТОРИЧНИХ ДОСЛІДЖЕННЯХ**

м. Кропивницький – 2025 рік

УДК 519.2:(93/94)(001.8)

Пасічник Н.О., Ріжняк Р.Я. Математико-статистичні методи в історичних дослідженнях. Навчальний посібник для здобувачів вищої освіти освітньо-наукового рівня галузі знань 03 Гуманітарні науки. Кропивницький, ЦДУ імені В. Винниченка, 2025. – 258 с.

Рецензенти:

Коцур Віктор Петрович, доктор історичних наук, професор, Заслужений працівник освіти України, дійсний член Національної академії педагогічних наук України, Університет Григорія Сковороди в Переяславі.

Орлик Світлана Владиславівна, доктор історичних наук, професор, професор кафедри історії, археології, інформаційної та архівної справи, Центральноукраїнський національний технічний університет

Посібник присвячений застосуванню сучасних кількісних методів для аналізу історичних даних. У ньому висвітлюються основні математичні та статистичні підходи, які дозволяють ефективно опрацьовувати великі обсяги інформації, заповнювати прогалини у джерелах та робити обґрунтовані висновки про минулі події й процеси.

Матеріал посібника охоплює методи описової статистики для аналізу історичних даних, побудову часових рядів та інтерпретацію їхніх трендів, використання регресійного аналізу для виявлення зв'язків між змінними, ймовірнісні моделі для реконструкції невідомих даних, основи використання статистичного програмного забезпечення.

Особливу увагу приділено практичним прикладам застосування кількісних методів в гуманітаристиці. Посібник також містить рекомендації щодо застосування сучасних цифрових технологій, включаючи різні спеціалізовані пакети.

Навчальний посібник буде корисним для істориків та здобувачів вищої освіти галузі знань 03 Гуманітарні науки, а також для студентів, викладачів та дослідників, які цікавляться застосуванням прикладних методів в гуманітарних дослідженнях.

Розглянуто на засіданні кафедри історії України та всесвітньої історії,
протокол № 7 від 16 грудня 2024 року

Затверджено до друку вченою радою
Центральноукраїнського державного університету
імені Володимира Винниченка
(протокол № 9 від «10» лютого 2025 р.)

©Н.О. Пасічник, Р.Я. Ріжняк

ЗМІСТ

ВСТУП: СПОЧАТКУ БУЛА КЛІОМЕТРІЯ...	5
РОЗДІЛ 1. МЕТОДОЛОГІЧНІ ЗАСАДИ ВИКОРИСТАННЯ МАТЕМАТИЧНИХ МЕТОДІВ В ІСТОРИЧНИХ ДОСЛІДЖЕННЯХ	9
1.1. Математизація наукового знання. Місце математичних методів в історичних дослідженнях	9
1.2. Математико-статистична обробка даних історичних джерел та математичне моделювання історичних процесів	11
1.3. Основні методологічні засади використання математичних методів	14
1.4. Приклади історичних проблем, вирішених за допомогою математичних методів	16
РОЗДІЛ 2. КОНТЕНТ-АНАЛІЗ ІСТОРИЧНИХ ТЕКСТІВ	26
2.1. Поняття про контент-аналіз. Принципи та особливості використання контент-аналізу в історичних дослідженнях	26
2.2. Приклад використання контент-аналізу для проведення дослідження на тему «Відображення особливостей освітнього, наукового й соціального середовища в текстах місій закладів вищої освіти України»	36
РОЗДІЛ 3. ДЕСКРИПТИВНА СТАТИСТИКА. ВИБІРКОВИЙ МЕТОД	43
3.1. Теоретичні відомості: описова статистика, вибірки, групування	43
3.2. Застосування інструменту EXCEL Описова статистика даних вибірки	58
РОЗДІЛ 4. АНАЛІЗ ВЗАЄМОЗВ'ЯЗКІВ	70
4.1. Принципи функціональної залежності. Метод найменших квадратів. Кореляція. Регресія. Детермінація. Рангова кореляція	70
4.2. Застосування інструментів EXCEL для аналізу взаємозв'язків	80

РОЗДІЛ 5. АНАЛІЗ ДИНАМІЧНИХ РЯДІВ.....	91
5.1. Основні характеристики динамічного ряду. Склад динамічного ряду: тренд, сезонна та стохастична компоненти. Вирівнювання. Стандартизація та нормалізація рядів. Кореляція. Автокореляція.....	91
5.2. Застосування інструментів EXCEL для аналізу динамічних рядів	102
РОЗДІЛ 6. ВИМІРЮВАННЯ СОЦІАЛЬНОЇ НЕРІВНОСТІ.....	115
6.1. Вимірювання в демографічній та соціальній сферах. Проблема вимірювання соціальної нерівності. Прості та складні гендерні індекси.....	115
6.2. Застосування інструментів EXCEL для аналізу простих та складних гендерних індексів	133
РОЗДІЛ 7. РЕТРОПРОГНОЗУВАННЯ	148
7.1. Методологічні засади ретропрогнозування. Типи імітаційно-прогностичних моделей: альтернативна та контрфактична. Будування динамічних рядів за неповними даними.	148
7.2. Застосування інструментів EXCEL для ретропрогнозування.....	200
РОЗДІЛ 8. МЕТОДИ БАГАТОМІРНОГО СТАТИСТИЧНОГО АНАЛІЗУ В ІСТОРИЧНИХ ДОСЛІДЖЕННЯХ.....	208
8.1. Принципи багатомірного статистичного аналізу. Кластерний, дискримінантний, факторний аналіз. Багатомірна класифікація.	208
8.2. Застосування інструментів EXCEL для багатомірного статистичного аналізу	243
ЗАКЛЮЧЕННЯ.....	251
ЛІТЕРАТУРА	254

ВСТУП:

СПОЧАТКУ БУЛА КЛІОМЕТРІЯ...

Кліометрія (від англ. *cliometrics*) – це напрям в економічній історії, що використовує кількісні методи, математичні моделі та економічні теорії для аналізу історичних даних. Назва походить від імені *Кліо*, музи історії в давньогрецькій міфології.

Кліометрія має такі основні характеристики. Вона використовує статистику та економетричні методи для аналізу історичних даних (наприклад, цін, зарплат, демографічних показників тощо). Теоретичною основою кліометрії є економічні моделі, які допомагають пояснювати поведінку людей або інститутів у минулому. Вчені, що займаються дослідженнями в галузі кліометрії часто працюють із великими обсягами даних, щоб знаходити закономірності та тенденції в історії. Загалом, це поле об'єднує методи економіки, історії та інколи навіть соціології чи політичних наук.

Кліометричні дослідження почали активно розвиватися у 1960-х роках. Вони змогли запропонувати нові підходи до аналізу економічної історії, особливо у таких сферах, як індустріалізація, рабство, економічне зростання та нерівність. Вперше термін кліометрія з'явився у пресі в грудні 1960 р. у статті Дж. Хьюгса, Л. Девіса і С. Рейтера «Аспекти квантитативного дослідження в економічній історії». Саме у 60-х рр. ХХ ст. ця галузь досліджень набула широкої популярності. Вирішальну роль тут відіграли прибічники кліометричного підходу Дуглас Норт і Уільям Паркер, які у 1960 р. стали редакторами «*Journal of Economic History*».

В історії виділяють 3 основні етапи та регіони поширення кліометрії:

1. США (1960–1970 рр.). Виникнення кліометрії пов'язане із США, де Роберт Фогель і Дуглас Норт використовували економетричні методи для вивчення історії залізничного транспорту та інституту рабства. Їхні дослідження викликали значний резонанс і започаткували дискусію щодо використання кількісного підходу в історичних дослідженнях. У 1960-х та 1970-х роках цей підхід швидко набув популярності в економічних та історичних колах США.

2. Європа (1970–1990 рр.). Кліометрія почала поширюватися в Західній Європі, зокрема у Великій Британії та Франції (в межах школи «Анналів»). Європейські дослідники використовували кліометричні методи для вивчення промислової революції, економічного зростання та соціальних структур. У Німеччині (Центр історико-соціальних досліджень Кельнського університету), Італії та інших країнах також виникли дослідницькі групи, що підтримували цей підхід.

3. Інші регіони (1990-ті роки – сьогодення). Кліометрія поступово проникала в інші країни світу, зокрема в Латинську Америку, Японію та інші регіони. У країнах, що розвиваються, кліометричні методи стали корисними для аналізу економічного розвитку, нерівності, трудової міграції та демографічних змін.

У 1993 р. Роберт Фогель и Дуглас Норт отримали Нобелівську премію з економіки за цикл робіт у галузі кліометрії. У рішенні Нобелівського комітету відмічалось, що премія присуджена «за розвиток нових підходів у дослідженнях з економічної історії, заснованих на застосування економічної теорії і кількісних методів для пояснення економічних і інституціональних змін».

Інституалізація кліометрії реалізувалася у відкритті кафедр економічної історії та створенні спеціалізованих видань. Проведення міжнародних конференцій сприяли налагодженню співпраці та взаємодії між дослідниками – міжнародні заходи, наукові видання та співпраця між університетами поширювали

кліометричні ідеї за межі США. Сприяв розвитку кліометрії і технічний прогрес – комп'ютеризація та доступ до великих баз даних значно спростили аналіз історичних даних, що дало поштовх для застосування кліометричних методів у більш широкому плані для проведення історичних досліджень (квантитативна історія).

З часом розвиток кліометрії («нової економічної історії») проходив у руслі більш ширшого напрямку – квантитативної історії, що включав застосування математичних методів у різних галузях історичного знання. До загалу теоретичних надбань української школи квантитативної історії відносять розробки інформаційних аспектів джерелознавства (науково-дослідна лабораторія комп'ютерних технологій історичних досліджень у Дніпропетровському університеті на чолі з В. Підгаєвським), концепцій і методів аналізу масових джерел (вивчення писемних джерел і рукописних книг (кодексів), історико-біографічних і бібліографічних джерел під час створення електронних ресурсів були проведені наприкінці 80 – поч. 90-х рр. ХХ ст. у Національній бібліотеці України імені В.І. Вернадського (Л. Дубровіна, В. Ясулайтіс, В. Чишко, О. Яценко, В. Попик, Л. Костенко, В. Омельчук, О. Іванова) та в Інституті українознавства імені І. Крип'якевича НАН України (Я. Ісаєвич); щодо музичних джерел аналогічні дослідження проводилися у Львівській консерваторії (Ю. Ясиновський)), методології застосування багатовимірного статистичного аналізу і математичного моделювання в історичних дослідженнях (Міждисциплінарний науковий центр прикладних досліджень ЦДУ ім. В. Винниченка).

Підводячи підсумок можемо навести висновки, зроблені відомим американським історіографом Г. Іггерсом («Історіографія у ХХ столітті: від наукової об'єктивності до постмодерністського виклику» 1997 р.): «У цілому модерністські тенденції мали в історичній науці цілий спектр проявів – від квантитативних підходів соціології і економіки, структуралізму школи «Анналів» до марксистського класового аналізу; усі ці течії орієнтувалися на модель наукового дослідження, визнану у природничих науках. На

відміну від традиційної історіографії з її увагою до індивідуалізму, вказані течії підкреслювали важливість вивчення соціальних структур і процесів соціальних змін».

Завдання даного навчального посібника зводиться до ознайомлення аспірантів з тими математичними методами, які вони можуть застосувати у своїх навчальних дослідженнях, на рівні наукових робіт без спеціальної математичної підготовки, без залучення складної електронно-обчислювальної техніки та які допоможуть молодим дослідникам у майбутній професійній діяльності.

РОЗДІЛ 1.

МЕТОДОЛОГІЧНІ ЗАСАДИ ВИКОРИСТАННЯ МАТЕМАТИЧНИХ МЕТОДІВ В ІСТОРИЧНИХ ДОСЛІДЖЕННЯХ

1.1. Математизація наукового знання.

Місце математичних методів в історичних дослідженнях

Процес наукового пізнання складається з трьох компонентів – методології, методики та техніки. Під методологією розуміється сукупність основних уявлень та ідей, принципів та прийомів пізнання, які є теорією методу. Шляхи та способи їх реалізації, набір відповідних правил та процедур складають методичку дослідження. Для будь-якого дослідження необхідні знаряддя, інструменти, що утворюють техніку. Між цими складовими існує тісний зв'язок, тобто активну роль може грати кожна з перерахованих тут частин. Водночас вони настільки взаємопов'язані, що їхнє самостійне існування у відриві один від одного неможливе, й всі вони підпорядковані головній меті – поглибленню та розширенню наших знань.

Сучасний стан історичної науки характеризується значним розширенням проблематики, пов'язаним із необхідністю, з одного

боку, узагальнити накопичений досвід і вийти на рівень фундаментальних робіт, які мають теоретико-концептуальний характер. Наприклад, вимагає комплексного підходу проблема ідентифікації українського історичного наративу. Подібне дослідження вимагає аналізу та узагальнення величезного обсягу джерел, різних за характером та формами вираження. З іншого боку, здобуття Україною незалежності відкрило можливість звернутися до багатьох, раніше заборонених тем, розширило джерельну базу дослідника, знявши гриф секретності з ряду архівних і бібліотечних комплексів. Це диктує потребу в детальному вивченні певних фактів, явищ, процесів. Крім того, низку історичних подій треба переосмислити, знявши з їхнього аналізу ідеологічні та імперські догми. Історія потребує підвищення об'єктивності своїх висновків та спостережень, підвищення точності.

Певну допомогу історика може надати математика. Під математикою зазвичай розуміється комплекс математичних дисциплін та наукових напрямів, що займаються вивченням абстрактних структур та операціями над об'єктами загальної природи, а значить і кількісними характеристиками соціальних явищ. В основі сучасних математико-статистичних теорій лежить поняття ймовірності. Під нею розуміється об'єктивна категорія, яка виступає мірою можливості того чи іншого результату, що характеризує з кількісною визначеністю можливість появи цієї події. За класичним визначенням ймовірність – це величина, рівна відношенню числа можливих випадків, що сприяють даній події, до всіх рівноможливих випадків. Припустимо, що у студентській олімпіаді беруть участь 50 осіб, із них 8 – студенти ЦДУ ім. В. Винниченка. У даному прикладі 50 – величина, що характеризує рівноможливі шанси до перемоги, а 8 – шанси перемоги студентів нашого університету. Отже, у 8 випадках із 50 можливих можуть перемогти студенти ЦДУ ім. В. Винниченка; чи $8:50 = 0,16$, тобто, ймовірність перемоги наших студентів дорівнює 0,16 (або 16 %).

1.2. Математико-статистична обробка даних історичних джерел та математичне моделювання історичних процесів

Чи піддаються соціальні явища ймовірнісному (з математичної точки зору) опису? Для ймовірнісних подій необхідне виконання низки умов:

1. Спостережувані явища або можуть бути повторені необмежену кількість разів, або одночасно здійснено спостереження за однаковими подіями у великій кількості. Не треба зайвий раз доводити, що експеримент, а отже, незліченне повторення подій в історії неможливе. Однак здійснити спостереження за великою кількістю однакових подій можна щодо масових джерел, масових сукупностей однорідних (однотипних за структурою) документів.

2. Щодо історії не можна говорити про незалежність історичних фактів, між ними існує причинно-наслідковий зв'язок, але в цьому випадку йдеться про незалежність документів. Кожен із них повинен формуватися самостійно, а не списуватися один з одного.

3. Наявність постійних умов під час створення джерельної бази обумовлює застосування математико-статистичних методів. Уникнення ідеї суворої детермінованості, обов'язковості історичних подій, введення в науковий обіг комплексів масових джерел дозволяє відносити явища історії до ймовірнісних, а отже розширити методичний арсенал введенням у нього математичних методів.

Основним завданням вивчення історичних явищ та процесів виступає розкриття їхнього внутрішнього механізму та всебічне пояснення їхньої сутності.

Кінцева мета будь-якого історичного дослідження полягає у виявленні закономірностей. Одні проявляються у поодиноких випадках (динамічні закономірності). При цьому характер динамічної закономірності встановлює поведінку кожної ознаки.

Інші – лише у масових випадках, тобто – у групі явищ, яка поруч із ознаками, властивими індивідуальним явищам, характеризується і спільними для всіх явищ ознаками (статистичними закономірностями).

Суспільне явище складається з маси індивідуальних і виявити історичну закономірність – значить знайти повторюваність всередині всієї маси явищ, де поряд з головними діє і безліч другорядних, нестійких, випадкових факторів. Це призводить до того, що в суспільстві немає певних динамічних закономірностей.

Використання в історичному дослідженні методів вивчення статистичних закономірностей дозволяє в масі випадкових факторів виділити основні, головні тенденції, притаманні в цілому явищу, що розглядається. Разом з тим не можна відкидати, випускати з поля зору й другорядні, малозначущі, а часом випадкові фактори, що з'являються і викликають ті чи інші стрибки в основній лінії розвитку суспільства.

Статистичні закономірності теоретично базуються на законі великих чисел, суть якого в найзагальнішому вигляді полягає в тому, що тільки при великій кількості спостережень формуються і виявляються багато об'єктивних закономірностей суспільних явищ. Вплив випадкових чинників, випадкових ознак стає меншим, якщо більше розглянуто поодиноких явищ. Так, наприклад, серед студентів першого курсу можна зустріти людину віком 28 років. Чи є це закономірним? Статистичне обстеження лише одного закладу вищої освіти показало, що середній вік першокурсника коливається в межах 18–20 років, те саме обстеження в рамках міста дає вік – 19 років. Отже, 28-річний студент на 1 курсі – явище випадкове, воно «загубилося» в масі спостережень. Однак, якби ми розглянули середній вік на основі вивчення всього 3-х студентів – 18, 20 та 28 років, то наша середня величина була б 22 роки. Тут значною мірою позначився вплив такого випадкового чинника, як 28-річний вік першокурсника.

Закон великих чисел означає, що випадкові відхилення, властиві одиничним явищам, у великій масі не впливають на

середній рівень досліджуваної сукупності. Відхилення індивідуальних елементів як би врівноважуються, нівелюються в масі явищ одного типу та перестають залежати від випадковостей. Саме ця властивість дозволяє вийти на рівень статистичної визначеності, статистичної закономірності. У законі великих чисел знайшов своє вираження зв'язок між необхідним і випадковим.

Статистична закономірність є кількісним вираженням певної тенденції, але не будь-яка статистична закономірність має історичний сенс. Можна виявити статистичну закономірність поширення куріння тютюну на території Наддніпрянщини в роки існування Запорізької Січі. Проте виявиться досить сумнівний вплив цієї тенденції на перебіг історичних подій. Аналізуючи отримані дані, історик з урахуванням змістовного, якісного підходу вирішує, чи характеризує знайдена статистична закономірність історичне явище, який ступінь узагальнення при цьому отримується, які умови його визначили тощо.

Таким чином, йдеться не про набуття історією математичної точності, а про розширення методичного арсеналу історика, про можливість отримання нових відомостей на більш досконалому кількісному та якісному рівні. Історична наука не втрачає своєї специфіки, так як математичні прийоми не замінюють якісний аналіз і не торкаються предмету історичної науки.

На сьогоднішній день не вироблено математичних методик, не пов'язаних з якісною стороною роботи. Не існує універсальних прийомів дослідження всіх історичних проблем, всіх історичних джерел. Вихідні теоретико-методологічні принципи історичної науки визначають цілі, шляхи та методи дослідження. На їх основі відбувається відбір, аналіз та узагальнення фактичного матеріалу.

1.3. Основні методологічні засади використання математичних методів

Процес дослідження співвідношення кількісного та якісного аналізу проходить протягом чотирьох етапів.

На першому етапі відбувається постановка проблеми, вибір джерел та визначення суттєвих ознак – це проходить при переважанні змістовного, якісного аналізу. Цей етап дуже важливий для подальшої роботи, так як від правильного виявлення важливих ознак залежить вибір методів аналізу. Тут відбувається певна формалізація джерела. Усі ознаки за своєю природою поділяються на кількісні (які виражаються числом) і якісні (визначаються словесно). Кількісні ознаки розкривають міру певних властивостей об'єкта, а якісні (атрибутивні) – наявність цих властивостей та їхню порівняльну інтенсивність. Різновидом якісних ознак виступають альтернативні, тобто такі, що приймають лише два значення (класичним прикладом якісної альтернативної ознаки є «стать» – або чоловіча, або жіноча).

Велика роль математики при розв'язанні задач, які пов'язані з підвищенням інформативної ваги джерел. Сучасники, фіксуючи ті чи інші аспекти історичних явищ, переслідують мету, відмінну від дослідницької. В зв'язку з цим дослідник не завжди може знайти у документах прямі відомості про цікаві аспекти явища. Практично будь-яке джерело містить приховану інформацію, що характеризує різноманітні взаємозв'язки, властиві історичним явищам. Вона виявляється внаслідок спеціальної обробки та аналізу даних.

На другому етапі обираються математичні методи залежно від структури джерела, характеру даних та сутності методів. Таке визначення проводиться у нерозривній єдності якісного та кількісного аналізу.

На третьому етапі спостерігається відносна самостійність кількісного аналізу. Відбувається з'ясування чисельних розподілів значень ознак, кількісних показників міри залежності між ними, визначаються показники інтенсивності впливу групи факторів на

систему, що вивчається, і т.п. Йде розрахунок показників за формулами (або з використанням статистичних програм). Усі явища без винятків характеризуються єдністю кількості та якості. Сутність того чи іншого явища, що виражає його якісну визначеність, буде розкрито лише тоді, коли буде виявлено кількісна міра даної якості.

На четвертому етапі проводиться змістовна інтерпретація отриманих результатів та побудова на їх основі теоретичних висновків. Ці операції вимагають від дослідника знання предмета, його кількісної та якісної сторони. Спільної схеми для такої інтерпретації не вироблено. Тут необхідно враховувати математичний аспект інтерпретації показників, отриманих у результаті розрахунків, з сутності застосованого методу. У той же час не можна забувати змістовний зміст проблеми, відступати від історичної можливості й реальності отриманих показників.

Між позначеними тут етапами існує найтісніший взаємозв'язок. Кожен попередній етап впливає наступний і навпаки. Наприклад, характер джерела визначає методіку його аналізу, водночас сам метод впливає на вибір ознак. Зазначена вище єдність якісних і кількісних характеристик явища має важливе значення при використанні математичних методів та інтерпретації їх результатів. Зміна кількісних параметрів може відбуватися в рамках однієї якості, а може приводити до набуття явищем нової сутності, нової якості.

Так, наприклад, збільшення значень такого кількісного показника, як розмір землекористування, досягнувши певного рівня, призводить до зміни соціального статусу селянина (від бідняка до середняка, від середняка до заможного селянина), тобто до появи нової якості.

Відмінність значень ознаки в різних одиниць сукупності у той самий період називається у статистиці варіацією. Вона є необхідною умовою існування та розвитку масових явищ. У житті кожної масової сукупності масовому процесу властива специфічна

міра варіації елементів, у яких цей процес протікає нормально, не змінюючи своєї якісної сутності.

1.4. Приклади історичних проблем, вирішених за допомогою математичних методів

Сьогодні історична наука досить широко користується математико-статистичними прийомами, чому значною мірою сприяє комп'ютеризація робочого місця дослідника. У зв'язку з цим найактуальнішими вважаються дві проблеми – розширення математичного інструментарію за рахунок впровадження в історіографію методів математичної логіки, теорії інформації, теорії графів тощо. Друга проблема – зберігання історичної інформації за допомогою інформаційних технологій, проблема створення баз і банків даних інформації з певних історичних тем, періодів, регіонів.

Розглянемо кілька прикладів проведених наукових пошуків, які були успішно реалізовані в межах діяльності університетського Міждисциплінарного наукового центру прикладних досліджень.

1. Аналіз забезпечення комп'ютерною технікою технічних університетів України протягом 1991–2011 років (детальніше – [28]).

Зміст дослідження та висновки. На основі економіко-статистичного аналізу часових рядів, що характеризують динаміку комплектування комп'ютерною технікою технічних вищих навчальних закладів України, визначені основні закономірності розвитку комп'ютерного забезпечення інформатизації технічних університетів України в економічних умовах, що були характерними для нашої держави протягом 1991 – 2011 років. Аналіз таблиці показників динаміки забезпеченості КТ технічних університетів України у розрахунку на 100 студентів денної форми навчання протягом вказаного періоду дає можливість стверджувати таке.

А. Протягом вказаного періоду середні показники забезпеченості технічних університетів КТ на 100 студентів зростали з лінійним коефіцієнтом 0,65. При цьому певне зниження та стабілізація показників забезпеченості в середині 90-х років може бути пояснене загальною макроекономічною ситуацією в Україні на той період. Падіння цих же показників на межі 2004 та 2005 років пояснюється протекціоністськими заходами, якими МОН України стимулювало виші до своєчасного оновлення технічних характеристик КТ.

Б. Мінливість показника забезпеченості КТ між різними технічними університетами мала до середини 90-х років тенденцію на різке зниження (з 84 % на початку 90-х до 32 % на середину десятиріччя); у подальшому мінливість показників коливалася у межах $40 \% \pm 10 \%$. При тому, що мінливість показників протягом всього періоду і мала тенденцію до зменшення, але залишалася вищою від середніх показників допустимої у статистиці варіативності динаміки процесу. Це яскраво свідчило про те, що різні технічні університети з різними успіхами виходили з непростой загальної макроекономічної ситуації в Україні протягом усього досліджуваного нами періоду.

В. Побудований нами з метою виявлення закономірностей варіювання забезпеченості КТ технічних університетів інтервальний статистичний ряд розподілу не мав властивостей нормальності через показники двох ВНЗ (ТДТУ та ДДМА), які протягом аналізованого періоду мали далеко вищі середні показники комплектації КТ, ніж інші технічні університети України. Але, як показав у подальшому дослідженні кластерний аналіз, показники динаміки розвитку забезпеченості КТ цих ВНЗ мало відрізнялися від показників динаміки основної групи університетів та академій.

Г. Аналіз коефіцієнтів мінливості забезпеченості КТ кожного з технічних університетів протягом 1991–2011 років показав високу варіативність результатів для різних ВНЗ, що пояснюється застосуванням різноманітних підходів керівництвом технічних університетів до організації менеджменту щодо визначення відповідного технічного рівня та оптимально необхідної кількості ПК навчального призначення виходячи з ліцензійних та акредитаційних вимог МОН

України, а також з власних фінансових можливостей та власних планів щодо розвитку інформаційних технологій у своєму ВНЗ.

Д. Визначення наявності множинного кореляційного зв'язку між часовими рядами, які характеризують динаміку комплектації технічних університетів комп'ютерною технікою протягом досліджуваного періоду показало, що наявним є порівняно тісний зв'язок між тенденціями розвитку такої забезпеченості різних ВНЗ. Це може бути пояснене, по-перше, явним існуванням елементів координації та структурних елементів системи моніторингу діяльності вишів з боку МОН України в області комплектування ВНЗ засобами комп'ютерної техніки протягом досліджуваного періоду; по-друге, наявністю тісних зв'язків на різних рівнях (як офіційних, так і неформальних) між головними менеджерами та відповідними керівниками структурних підрозділів різних технічних університетів України.

2. Контент-аналіз нумізматичних досліджень українських історичних періодичних видань (1990-2016 роки) (детальніше – [29]).

Зміст дослідження та висновки. У розвідці наводяться результати дослідження щодо використання можливості кількісного контент-аналізу для вивчення змісту контенту українських історичних журналів на предмет виявлення закономірностей висвітлення в них результатів наукових досліджень зі спеціальної історичної дисципліни нумізматики – визначені смислові категорії контент-аналізу видань, обсяг вибірки, одиниці контексту та одиниці обрахунку, проведений кількісний та якісний аналіз журнального контенту. Сформульовані такі висновки.

А. Для виявлення направленості та активності висвітлення результатів нумізматичних досліджень були відібрані чотири наукові історичні журнали – «Спеціальні історичні дисципліни: питання теорії та методики» (33 номери журналу з 1997 року), «Наукові записки з української історії» (19 номерів видання з 2007 року), «Вісник Харківського національного університету імені В.Н. Каразіна. Серія: Історія» (12 номерів видання з 2007 року), «Український історичний журнал» (180 номерів журналу з 1990 року). Вибір проводився не

випадковим чином, а виходячи з того, що названі видання містять більш, ніж 96 % контенту результатів нумізматичних досліджень.

Б. Найбільш популярними дослідженнями, результати яких публікувалися в обраних для аналізу виданнях, стали наукові пошуки, присвячені проблемам нумізматики нового часу (майже половина від загального обсягу досліджень). Майже чверть від загального обсягу досліджень були присвячені проблемам середньовічної нумізматики. Інші п'ять категорій досліджень ледь набрали третину загального обсягу. Це можна пояснити наявністю багатого нумізматичного матеріалу епохи середньовіччя та нового часу, використанням дослідниками детально розроблених довідкових матеріалів та каталогів монет цих періодів, загальною розробленістю цієї тематики в науковій літературі.

В. Фаховість видання «Спеціальні історичні дисципліни: питання теорії та методики» забезпечила йому найбільший обсяг контенту з поміж інших журналів. Хоча, за «питомою» вагою обсягу публікацій на своїх сторінках видання незначно поступається виданню «Наукові записки з української історії», яке і зайняло «друге місце» за загальним обсягом нумізматичного контенту. Це закономірно, так як саме на базі Переяслав-Хмельницького педагогічного університету ім. Г.С. Сковороди проводяться щорічні науково-практичні конференції з проблем нумізматики та створено науковий центр з організації нумізматичних досліджень. Два інші видання розкривають у своїх публікаціях загальні історичні проблеми – тому не мають вагомих причин для концентрування уваги лише на проблемах нумізматики.

Г. Аналіз розподілу обсягу контенту за окремими темами досліджень між названими науковими виданнями дав можливість визначити «лідерів» за окремими категоріями. Так, «лідерство» у висвітленні проблем античної нумізматики у видання «Вісник Харківського національного університету імені В.Н. Каразіна. Серія: Історія» пояснюється людським фактором – наявністю у Харківському національному університеті фахівця східної нумізматики високої кваліфікації К.В. Мизгіна, а «лідерство» видання «Наукові записки з української історії» у висвітленні проблем нумізматики Візантійської імперії та християнського сходу – участю фахівця з даної проблематики

професора В.М. Орлика у якості організатора та активного учасника вже згаданого центру вивчення нумізматики та наукових конференцій на базі Переяслав-Хмельницького педагогічного вишу. Причиною найбільшого обсягу публікацій з проблематики середньовічної нумізматики та нумізматики нового часу, що належить виданню «Спеціальні історичні дисципліни: питання теорії та методики» є наявність нумізматичного матеріалу, каталогів та довідників. Можемо лиш додати, що загальний обсяг висвітлення результатів досліджень саме за вказаними категоріями становить у даному виданні 44 % від загального обсягу публікування у цьому журналі всіх результатів нумізматичних досліджень.

3. Прогностичний аналіз публікаційної активності науковців університетів України у контексті їх входження до рейтингу QS World University Rankings (детальніше – [17]).

Зміст дослідження та висновки. QS World University Rankings є одним із найавторитетніших рейтингів, що показує ефективність закладу вищої освіти порівняно з іншими. Наукометричний показник цитування, який визначає п'яту частину підсумкового результату, є досить важливим у названому рейтингу і розглядається як показник якості викладацького складу. Ураховуючи це, основною проблемою статті є вивчення перспективи розвитку цитованості науково-педагогічних працівників українських університетів для забезпечення входження до вказаного рейтингу більшої кількості ЗВО нашої держави. Розв'язання окресленої проблеми актуалізується через фактичну відсутність детального наукового аналізу індексування та цитування праць українських науковців, що публікуються у визнаних Міністерством освіти і науки України наукометричних базах.

З метою проведення аналізу публікаційної активності вчених українських університетів використані математико-статистичні методи. Для побудови прогностичних моделей розвитку тенденцій публікаційної активності науково-педагогічних працівників вітчизняних ЗВО використовувався метод експоненційного згладжування за моделлю Хольта, яка показала найкращі показники узгодженості емпіричних даних та отриманих теоретичних моделей. Для аналізу публікаційної активності вчених університетів України у контексті врахування

методології побудови рейтингу QS проведено відбір університетів за авторською методикою.

У результаті дослідження вдалося визначити особливості динаміки розвитку публікаційної активності науково-педагогічних працівників українських університетів протягом останніх 9 років, головна з яких полягала у прискоренні темпів зростання цитування публікацій українських вчених з моменту переорієнтації на європейський досвід розвитку науки й освіти та, як наслідок, прийняття Кабінетом Міністрів України змін до умов провадження освітньої діяльності. Крім того, прогнозування показників цитування наукових праць учених ЗВО України на 2021–2023 роки, які враховуються у 5-й частині розрахунку рейтингу QS за його методикою (нормована кількість цитувань, отриманих статтями університету за 5 років за даними бази Scopus у розрахунку на одного викладача), показало, що для українських університетів недостатність цього показника є досить серйозною проблемою у контексті конкуренції з закордонними університетами. Для її розв'язання як на державному рівні, так і на рівні закладів вищої освіти України слід розробити заходи дієвої підтримки публікаційної активності університетської науки.

4. Публікаційна активність учених педагогічних університетів України (2010–2020 рр.) через призму баз даних Scopus та Web of Science (детальніше – [18]).

Зміст дослідження та висновки. Метою пошуку було визначення особливостей та динаміки наукової активності та публікаційної видимості вчених педагогічних закладів вищої освіти України у вимірах наукометричних баз даних Scopus та Web of Science протягом другого 10-річчя XXI століття.

Авторами була обрана генеральна сукупність українських педагогічних університетів, профілі яких були представлені принаймні в одній з баз даних – Scopus або Web of Science, та з використанням математико-статистичних методів проаналізовані загальна кількість проіндексованих публікацій, загальна кількість цитувань, h-індекс та структура основних напрямків наукового пошуку вчених з педагогічних університетів станом на кінець вересня 2020 року. Крім того, для

кожного педагогічного вишу по базі Web of Science були проаналізовані вищеназвані характеристики у вигляді часового ряду публікаційної активності закладу вищої освіти за останніх 10 років.

В результаті було виявлено, що між основними загальними показниками педагогічних університетів у названих наукометричних базах існує значущий зв'язок як між показниками всередині баз, так і між відповідними показниками з різних баз. Аналіз індексування обома базами досліджень вчених педагогічних університетів продемонстрував різне відношення менеджменту компаній-постачальників послуг до розширення охоплення публікацій вчених. Аналіз динамічних рядів показників публікаційної активності вчених з педагогічних ЗВО протягом 2010–2020 років показав, що з 2015 року кількість проіндексованих публікацій та показники індексування посилань за базою Web of Science у більшості закладів вищої освіти педагогічного профілю різко зросли через запровадження вимоги щодо наявності індексованих базами праць науковців як обов'язкової ліцензійної умови та початок індексування базою Web of Science ряду нових українських наукових видань з гуманітарних, педагогічних та міждисциплінарних проблем. Крім того було виявлено, що статті науковців педагогічних університетів, які були опубліковані протягом досліджуваного періоду, в різних ЗВО по-різному вплинули на формування ядра h -індексу відповідного вишу.

5. Biographical Materials of Mathematicians and Natural Scientists in “The Bulletin of Experimental Physics and Elementary Mathematics” (1886–1917): meaningful and content analysis (детальніше – [10]).

Зміст дослідження та висновки. Представлені результати дослідження особливостей біографічних та просопографічних матеріалів про відомих математиків та дослідників природи, які друкувалися в одному з найбільш авторитетних журналів «Вісник дослідної фізики та елементарної математики», який видавався в Києві та Одесі протягом 1886–1917 рр. Фактично журнал був неофіційним періодичним друкованим органом математичного відділення Новоросійського товариства дослідників природи. Мета дослідження полягала у проведенні змістовного та контентного аналізів статей журналу, в яких розкриваються біографії, особливості трудової діяльності та деталі

участі у наукових дослідженнях відомих вчених. При цьому авторами для проведення змістовного аналізу предмета дослідження використовувалися наукові методи – аналізу та синтезу, узагальнення і систематизації. У процесі проведення кількісного контентного аналізу використовувалися квантифікація тексту, збирання емпіричних даних, їх узагальнення та математико-статистична обробка.

В результаті проведеного дослідження біографічних матеріалів журналу за весь час його видання автори дійшли таких висновків.

По-перше, матеріали журналу про відомих вітчизняних і зарубіжних математиків і дослідників природи носили як біографічний, так і просопографічний характер. Причому мотивами появи таких публікацій, як правило, були «круглі» дати з моменту народження, початку творчої діяльності, з моменту виходу важливого твору або з моменту смерті (або факт смерті) суб'єкта публікації. Усі подібні статті в журналі умовно класифіковані авторами на короткі біографічні та просопографічні нотатки, розгорнуті біографічні та просопографічні статті, некрологи.

По-друге, загальний обсяг біографо-просопографічних матеріалів за період видання журналу становив 2,88 % від загального обсягу журналу. При цьому журнал містив матеріали про 84 вчених. Найбільший журнальний обсяг редакцією (більше 5 сторінок для кожного) був наданий 27-ми вченим і це склало 72,7 % від загального обсягу біографічного матеріалу; до цієї групи потрапили 9 вітчизняних учених та 18 іноземців. На кожного вітчизняного вченого у загальному розрахунку редакція виділяла в середньому 4,89 сторінки, а на кожного іноземного – 5,89 сторінки. Найбільший обсяг матеріалів був про фізиків (46,42% від загального обсягу біографічних матеріалів) та про математиків (35,44%).

По-третє, результати змістовного аналізу дозволяють зробити висновок про високу якість друкованих біографо-просопографічних матеріалів журналу. Цьому багато сприяв ретельний підбір авторів статей (чи джерел матеріалів), і навіть особливі вимоги редакції до документування джерел у посторінкових посиланнях.

По-четверте, «неуважність» редакції журналу до цілого ряду «круглих дат» знаменитих вітчизняних та зарубіжних математиків та фізиків, а також нерівномірні обсяги, що виділялися журналом для біографістики, свідчать про відсутність чіткої редакційної політики журналу щодо виділення друкованих обсягів для систематичних біографічних заяв та досліджень.

6. Е.К. Шпачинський – засновник науково-популярної журналістики кінця ХІХ ст. на теренах України (детальніше – [26]).

Зміст дослідження та висновки. Дослідження присвячене вивченню особливостей діяльності Е. Шпачинського як засновника науково-популярної журналістики на теренах сучасної України кінця ХІХ століття. У дослідженні використовуються описові та кількісні методи.

Для змістовного аналізу використовується аналіз інформації видання та синтез логічних посилань та висновків, а також узагальнення поданих у публікаціях фактів і систематизація описів подій з життя та редакторської діяльності Е. Шпачинського. У процесі проведення кількісного аналізу використовувалися квантифікація тексту, збирання емпіричних даних, їх узагальнення та математико-статистична обробка.

Дослідженням встановлено: А. Е. Шпачинському протягом своєї редакторської та видавничої діяльності (1886–1897 рр.) у Віснику вдалося утвердити існування науково-популярного фізико-математичного журналу на теренах сучасної України, згуртувавши навколо нього найкращі сили вчених-популяризаторів, творчих викладачів, цінителів фізико-математичних знань і реалізувавши в його програмі надійний зв'язок з потребами викладання фізики й математики при підготовці учнівської молоді.

Б. Е. Шпачинський у Віснику продовжив програму В. Єрмакова з елементарної математики, розширивши її відомостями з фізики. Поряд зі статтями про традиційні й відомі об'єкти фізики та математики, у Віснику друкувалася велика частка статей, що розкривали нові горизонти цих наук і давали інформацію про результати нових досліджень.

В. Е. Шпачинський спромігся за період своєї редакторської діяльності згуртувати широкі кола педагогічної громадськості до обговорення на сторінках Вісника найважливіших питань викладання математики й фізики.

Г. Е. Шпачинський самотійно вів активну та різносторонню редакторську діяльність Вісника як автор різноманітних дописів від редакції.

Д. Увесь період редакторської діяльності Е. Шпачинського був пов'язаний з постійною боротьбою із зовнішніми несприятливими умовами видання Вісника: постійна потреба в коштах, небажання чиновників від освіти надати підтримку існуванню журналу, брак часу для реалізації організаторської діяльності. Ці умови перешкоджали повному розкриттю таланту кваліфікованого фізика-експериментатора, популяризатора наукових знань та педагога-викладача, яким був Е. Шпачинський.

РОЗДІЛ 2.

КОНТЕНТ-АНАЛІЗ ІСТОРИЧНИХ ТЕКСТІВ

2.1. Поняття про контент-аналіз. Принципи та особливості використання контент-аналізу в історичних дослідженнях

Найбільш велику групу історичних джерел становлять розгорнуті індивідуальні тексти. Джерелознавство не відносить наративні джерела до масових, а значить вони не відповідають вимогам, що висуваються до ймовірнісних подій і до них не можуть бути застосовані математичні методи. Однак письмовий текст має статистичну структуру. Певні характеристики, що містяться в ньому, можуть бути описані за допомогою ймовірнісних законів. Таким чином, у разі потреби наративне джерело можна перетворити на масове шляхом частотних, класифікаційних перетворень, тобто методом контент-аналізу.

Контент-аналіз складається із двох основних розділів. У першому, класифікаційному розділі, з дослідницької мети та інформативного потенціалу джерел, виділяється сукупність ознак, які часто зустрічаються в документах. Тут відбувається формалізація документа, виробляється якась подібність до анкети. Деякі дослідники розглядають процедуру контент-аналізу, поділяючи перший класифікаційний розділ на дві дії – дію якісного та дію кількісного характеру.

Складність роботи на класифікаційному розділі полягає не скільки у виділенні ознак, як у визначенні їхньої градації та її ступенів. Це пов'язано зі специфікою мови, особливостями словесних характеристик, системою їхнього виміру. Наприклад, одна й та ж ознака може зустрічатися в тексті з вираженням різними словами-синонімами, алегоріями, розгорнутими описами, мати, залежно від контексту, різне емоційне та смислове забарвлення, набувати і позитивних, і негативних значень. Як смислова одиниця аналізу тексту може виступати тема, виражена в цілій статті, у фрагменті тексту, аж до окремих слів і навіть складів і літер.

В кінцевому результаті дослідник отримує набір ознак, які зустрічаються в документах, у текстах досить багато разів і які приймають змінні значення (в історіографії отримані первинні поняття називаються символами).

Далі вводяться категорії – загальніші, великі показники, що є класами символів. Наступним кроком виступає підрахунок частоти появи у тексті кожного символу та частоти їхніх смислових зв'язків. Якщо є необхідність, то визначається частота відношення автора (авторів) тексту до отриманих смислових одиниць.

Таким чином, можна сказати, що під час першого розділу проведення контент-аналізу вирішується питання «що рахувати?». Робота не може бути формалізована, вона визначається методологічними принципами дослідника, рівнем його професійної кваліфікації. У літературі у зв'язку з цим висловлено сумнів щодо об'єктивності контент-аналізу. Однак треба відзначити, що метод дозволяє простежити всі кроки дослідника, всі етапи його роботи й відтворити їх заново, ретельно перевірити результати. У цьому сенсі роботи, виконані з урахуванням контент-аналізу, вигідно відрізняються від традиційного підходу, коли історик розкриває свою концепцію, ілюструючи її окремими фрагментами з тексту джерела. Тут особиста позиція автора більшою мірою визначає характер результатів вивчення тексту, адже інший історик, звернувшись до того ж джерела може

отримати зовсім інші висновки та проілюструвати їх іншими фрагментами того ж тексту.

Під час реалізації другого розділу контент-аналізу розв'язується питання «як рахувати?». Залежно від характеру кількісних даних, від частотних класифікацій, результатів згрупувань визначається процедура розрахунку показників за вибраними математико-статистичними методами.

Вперше метод контент-аналізу був застосований 1910 року Максом Вебером з метою проведення кількісно-орієнтованого аналізу текстів друкованих засобів інформації, що висвітлювали політичні акції в Німеччині. У подальшому цей метод активно використовувався в історичних та політологічних дослідженнях саме з 50-х років ХХ століття [3]. На сьогодні досить популярним є один з варіантів кількісного контент-аналізу, що запропонований німецьким політологом Вернером Фрю. В. Фрю таким чином визначив кількісний контент-аналіз: емпіричний метод систематичного, інтерсуб'єктивно відтворюваного опису змістовних та формальних ознак текстів з метою отримання на цій основі висновків, що відносяться до позатекстової реальності [4].

Отже, контент-аналіз – це метод дослідження, що використовується для систематичного аналізу текстового, візуального або аудіо-візуального контенту з метою виявлення шаблонів, тенденцій, тематик або інших характеристик. Крім історії, цей метод може застосовуватися в різних дисциплінах, включаючи мас-медіа, соціологію, політичну науку, психологію, маркетинг та інші сфери. Основні етапи контент-аналізу включають:

1. Формулювання дослідницької проблеми: визначення конкретного аспекту, який досліджується, і встановлення мети дослідження.

2. Визначення вибірки контенту: вибір конкретних джерел або документів, які будуть аналізовані, і встановлення критеріїв вибору.

3. Розробка кодувальної схеми: створення системи категорій або кодів, які будуть використовуватися для класифікації або маркування вмісту.

4. Кодування контенту: аналіз і класифікація тексту, зображень або іншого контенту відповідно до розробленої кодувальної схеми.

5. Аналіз даних: обробка та аналіз закодованого контенту для виявлення шаблонів, тенденцій або інших характеристик.

6. Інтерпретація результатів: висновки на основі аналізу даних і їхня інтерпретація в контексті дослідницької проблеми.

Принципи контент-аналізу визначають основні принципи й підходи, які використовуються при проведенні аналізу контенту.

1. Систематичність: контент-аналіз повинен бути проведений систематично, щоб гарантувати об'єктивність та повноту аналізу; це включає установа чітких критеріїв вибірки, розробку однорідної кодувальної схеми та здійснення аналізу відповідно до цих критеріїв.

2. Репрезентативність: важливо, щоб вибірка контенту була репрезентативною для досліджуваного явища або популяції; це допомагає забезпечити достовірність та генералізованість результатів аналізу.

3. Об'єктивність: аналіз контенту повинен проводитися об'єктивно, без упередженості або впливу дослідника; це може вимагати розробки чітких і однозначних критеріїв кодування та використання незалежних оцінювачів.

4. Надійність і відтворюваність: результати контент-аналізу повинні бути надійними й відтворюваними; це означає, що аналіз повинен бути здійснений таким чином, щоб інші дослідники змогли повторити або перевірити результати.

5. Контекстуалізація: важливо враховувати контекст досліджуваного контенту при аналізі; це означає врахування історичного, культурного, соціального та політичного контексту, щоб зрозуміти значення та вплив контенту.

6. Достовірність: контент-аналіз має базуватися на достовірних даних; це включає узгодженість та достовірність використаних джерел, а також обґрунтованість критеріїв кодування.

7. Прозорість і відкритість: важливо, щоб методологія та процес аналізу були прозорими і відкритими для перевірки й реплікації іншими дослідниками.

Викладені принципи допомагають забезпечити якість і надійність проведення контент-аналізу та достовірність його результатів.

Понятійний апарат контент-аналізу включає ряд термінів та концепцій, які використовуються для опису та аналізу текстового, візуального або аудіо-візуального контенту. Основні поняття, які часто зустрічаються в контент-аналізі, є такими:

1. Кодування – процес присвоєння кодів або категорій до конкретних фрагментів контенту з метою подальшого аналізу. Це можуть бути числові коди, текстові мітки або інші символи, що відображають певні аспекти контенту.

2. Кодувальна схема – система категорій або кодів, що використовується для класифікації контенту. Кодувальна схема визначає основні аспекти аналізу та набір категорій, які будуть використовуватися під час кодування.

3. Юніт кодування – одиниця аналізу, до якої застосовуються коди або категорії. Це може бути слово, фраза, абзац, зображення, відеофрагмент або будь-який інший елемент контенту.

4. Інтеркодерна надійність – міра узгодженості між різними кодерами, які незалежно кодують той самий контент. Інтеркодерна надійність використовується для оцінки достовірності аналізу.

5. Частота – кількість випадків, коли певний код або категорія зустрічаються в контенті. Частота може бути використана для визначення статистичної важливості певних аспектів контенту.

6. Тема – основна ідея або поняття, які виражені в контенті. Теми можуть бути ідентифіковані шляхом аналізу ключових слів, понять або ідей, які повторюються в тексті.

7. Сентимент – тон або емоційна оцінка, виражена в контенті. Сентимент може бути позитивним, негативним або нейтральним і може бути визначений шляхом аналізу вживаних мовних засобів та контексту.

8. Контентний аналізатор – людина або програмний засіб, який здійснює аналіз контенту, кодує його відповідно до кодувальної схеми та проводить інші операції з обробки даних.

Цей понятійний апарат допомагає стандартизувати та узагальнити аналіз контенту, роблячи його більш об'єктивним та систематичним.

Контент-аналіз може використовуватися для різноманітних цілей, включаючи вивчення вмісту мас-медіа, аналіз змісту політичних дебатів, дослідження стереотипів та переконань, оцінку ефективності маркетингових кампаній та інше. Цей метод дозволяє отримати об'єктивні дані про вміст, які можуть бути використані для подальшого дослідження та прийняття рішень.

Особливий інтерес представляє можливість використання контент-аналізу для виявлення та вивчення прихованої інформації, що міститься в джерелі. Звістка, що будь-яке історичне джерело, як наративне, так і статистичне, містить у собі, з одного боку, явну, виражену або пряму інформацію і, з другого боку, приховану, такою, що є пов'язаною або структурною. Явна інформація відбиває головну мету автора (авторів) документа. Вона фіксувалася авторами джерела свідомо, була близька й зрозуміла сучасникам, використовувалася ними за прямим призначенням.

Прихована інформація потрапляла в документ ненавмисно, поза суб'єктивною волею людей, причетних до збору та фіксування даних. Вона є результатом стихійного відображення у джерелі важливих особливостей часу через взаємозв'язки між процесами, явищами, властивостями об'єктів.

Прихована інформація має дві важливі особливості. По-перше, вона виражається опосередковано – через події, що містяться в явній інформації, через їхні характеристики. По-друге, маючи колосальний обсяг (так як взаємозв'язки, як правило, складні й

різноманітні), вона буває об'єктивнішою і достовірнішою через своє стихійне походження, ніж ті показники, які навмисне закарбовувалися в документі.

Іноді прихована інформація з'являється у тексті джерела свідомо. Це буває в тих випадках, коли автори навмисне маскують свої справжні цілі, ідеї, переконання, використовують езопову мову.

У ряді випадків частина інформації виявляється прихованою через несистематизованість і об'ємність первинних матеріалів. Текст деяких документів – промов, листівок, публікацій періодичного друку, анкет, щоденників, листів, мемуарів тощо – несе у собі багато неясного через суперечливість авторської позиції. У цьому випадку лише контент-аналіз визначить та кількісно висловить ступінь суперечливості та міру нестійкості ідейних позицій автора. Перевага контент-аналізу полягає в тому, що за його допомогою історик отримує можливість переведення масової текстової інформації в кількісні показники. Це значно знижує суб'єктивізм дослідження, розв'язує проблему його перевірки.

Таким чином, сильні сторони контент-аналізу включають:

1. Об'єктивність – контент-аналіз може бути виконаний об'єктивно і систематично за допомогою чітких методологічних підходів і кодувальних схем.

2. Масштабованість – цей метод може бути застосований до великих обсягів контенту, що дозволяє отримувати представлення про широкомасштабні тенденції, схеми та моделі.

3. Множинність джерел – контент може бути зібраний з різних джерел, включаючи текст, візуальні та аудіо-візуальні матеріали, що дозволяє отримати комплексне уявлення про явище.

4. Глибина аналізу – за допомогою контент-аналізу можна розкрити різні аспекти контенту, такі як теми, сентимент, структура, стилістика, а також зв'язки між концепціями та ідеями.

5. Широке застосування – контент-аналіз може бути застосований не лише в історичних дослідженнях, а також в різних галузях, таких як мас-медіа, соціологія, політична наука, психологія, маркетинг, дослідження здоров'я та багато інших.

6. Можливість порівняння – контент-аналіз дозволяє порівнювати різні джерела чи періоди часу для виявлення змін у змісті або тенденціях.

7. Дослідницька гнучкість – метод контент-аналізу може бути адаптований до різних дослідницьких завдань та цілей, що робить його універсальним і гнучким.

Ці сильні сторони роблять контент-аналіз ефективним інструментом для дослідження та аналізу різних типів контенту з різних джерел і в різних контекстах. Втім, хоча контент-аналіз має свої переваги, він також має свої обмеження та слабкі сторони:

1. Суб'єктивність кодування – хоча кодування може бути стандартизованим, існує ризик суб'єктивного тлумачення дослідником або кодером контенту, що може призвести до неточних результатів.

2. Обмеженість вибірки – результати контент-аналізу можуть бути обмежені обсягом та репрезентативністю вибірки, яка може не відображати повністю різноманітність контенту.

3. Часова та ресурсна витратність – проведення контент-аналізу може бути часово та ресурсно витратним процесом, особливо якщо аналізується великий обсяг контенту.

4. Неадекватність для складних аналізів – у деяких випадках контент-аналіз може бути недостатнім для аналізу складних або мульти-модальних даних, таких як відео або музика.

5. Нехтування контексту – аналіз контенту може не завжди враховувати контекстуальні чинники, які можуть впливати на інтерпретацію даних.

6. Підвищений ризик помилок при кодуванні – при кодуванні контенту існує ризик помилок, особливо якщо кодування виконується вручну (слід пам'ятати, що навіть незначні помилки можуть спотворити результати аналізу).

7. Можлива суб'єктивність при визначенні категорій – розробка кодувальної схеми може бути суб'єктивною і варіюватися в залежності від інтерпретації дослідника.

Ці обмеження слід враховувати при застосуванні контент-аналізу та інтерпретації його результатів.

Слід зазначити, що контент-аналіз досить успішно застосовується в історії для різноманітних наративних джерел. З його допомогою обробляється інформація періодичних видань, щоденників, мемуарів, листування, авторських текстів (філософські трактати, записки, різноманітні описи), літописів тощо. Ефективність застосування контент-аналізу пов'язані з можливістю реалізації системного підходу до аналізу змісту історичного джерела. Цей метод допомагає отримати більш об'єктивну, обґрунтовану аргументацію висновків, а в ряді випадків отримати й приховану в джерелі інформацію.

На основі математичних методів отримані корисні результати при дешифруванні текстів (наприклад, у разі шифрування або невідомої мови написання), атрибуції пам'ятника писемності, встановлення авторства. Використання теорії інформації у текстології дозволяє оцінити кількість проміжних текстів, що передують даному. За допомогою математичного інструментарію можна встановити оригінал серед численних текстів, визначити місце створення тексту, датувати його. Можлива реконструкція джерела, його раніше втрачених фрагментів, очищення оригінального тексту від пізніших нашарувань. Однак, незважаючи на накопичений досвід, безліч методологічних та методичних проблем, пов'язаних з виміром наративних джерел, залишаються невирішеними.

Сучасний етап використання математико-статистичних методів, зокрема аналізу текстових джерел, пов'язані з застосуванням комп'ютерної техніки. Ще на початку 60-х років з'явилася нова галузь інформатики – технологія баз даних. Новий етап використання ЕОМ в історичних дослідженнях настав у 80-ті роки разом із впровадженням у практику пошуків персональних комп'ютерів. Сьогодні історик виступає не лише як користувач історичних джерел, а й як творець нової історичної інформації. Розробляються методичні та методологічні проблеми формалізації

джерела, програмного забезпечення історичних досліджень, змісту та аналізу баз машинної інформації.

Наведемо приклади декількох програмних засобів, призначених для проведення контент-аналізу. Деякі з них спеціалізуються на аналізі текстового контенту, інші – на візуальному або аудіо-візуальному контенті.

1. NVivo – це програмне забезпечення для якісного та кількісного аналізу якісних даних, яке дозволяє аналізувати текстовий, візуальний, аудіо та інший контент.

2. MAXQDA – інструмент для якісного аналізу якісних даних, який дозволяє виконувати кодування, організацію та аналіз текстового, аудіо, відео та зображень.

3. Atlas.ti – програмне забезпечення для якісного аналізу даних, яке дозволяє аналізувати текстовий, візуальний, аудіо та інший контент, а також будувати та візуалізувати результати.

4. Dedoose – це інтернет-платформа для якісного та кількісного аналізу якісних даних, яка дозволяє спільноті дослідників працювати над аналізом.

5. Content Analysis Tool (CAT) – безкоштовне програмне забезпечення для аналізу тексту, яке дозволяє проводити різні види контент-аналізу, такі як аналіз тематик, семантичний аналіз та інші.

6. WordStat – це програмне забезпечення для аналізу тексту, яке дозволяє проводити кількісний та якісний аналіз текстового контенту, використовуючи різні методи.

Ці програмні засоби мають різні функції та можливості, і вибір конкретного інструменту залежить від потреб дослідження, типу аналізу, доступності та інших факторів.

Новою тенденцією, на яку має зважати сучасна історична наука, є здійснення діловодства та роботи низки офіційних та архівних установ у машинній формі. Архіви, бібліотеки, музеї, дослідні організації, статистичні бюро, центри документації створюють дедалі більше машинних файлів для збереження та вторинного використання інформації.

Незважаючи на те, що не всяке історичне джерело потребує переведення його в машинну форму й мало ще розроблено спеціальних програм, а стандартне програмне забезпечення не може повністю задовольнити історика, робота в цьому напрямі має продовжуватися і альтернативи в цьому питанні немає. Опанування математичними методами та комп'ютерними технологіями є свідченням професіоналізму сучасного історика.

2.2. Приклад використання контент-аналізу для проведення дослідження на тему «Відображення особливостей освітнього, наукового й соціального середовища в текстах місій закладів вищої освіти України»

Розглянемо результати проведеного нами в межах діяльності Міждисциплінарного наукового центру прикладних досліджень порівняльного контент-аналізу місій українських університетів, які були зафіксовані з відкритих джерел відповідних закладів вищої освіти станом на вересень 2021 року (детально з ходом та результатами дослідження можна ознайомитися з праці [16]).

Функції університету розглядаються як галузь діяльності, що відповідає його специфіці та соціальному призначенню. Чітке розуміння та визначення функцій забезпечує стійку діяльність університету на користь соціальної системи. Основні функції університету обґрунтували класики «ідеї університету» В. Гумбольдт, Д. Ньюмен, М. Вебер, К. Ясперс, Х. Ортега-і-Гассет та ін. Розглянемо впровадження класичних постулатів теоретиків «класичного університету» на прикладі закладів вищої освіти України педагогічного профілю.

Зазначимо, що однією з важливих особливостей сучасної ситуації є віддалення від загальних орієнтирів, вироблених у системі класичної університетської освіти. Класика пов'язана зі

стійкістю, традиціями, зразками та еталонами, з якими необхідно звіряти свою діяльність. У сучасній освіті на зміну традиціям прийшли інновації, на зміну стійкості – невизначеність, на зміну однаковості – плюралізм.

В даний час до пошуку «ідеї університету» включені не лише вчені та політики, а й самі університети. У текстах місій університети визначають своє призначення, розуміння своєї високої мети, прояснюють «чим є організація і якою вона прагне бути», показують «відмінність організації від інших, їй подібних».

Дослідники перспектив розвитку університетської освіти оцінюють процес активізації діяльності університетів з вироблення місій з двох крайніх позицій:

– як відповідь на необхідність виживати в умовах «усі зростаючої прагматики», спробу зберегти гуманітарну складову університетської освіти;

– як зростаючу прагматику університетів, «нову дискурсивну практику» сучасних університетів, де місія стає маркетинговим ходом для залучення нових абітурієнтів та ресурсів.

На початковому етапі контентного аналізу використовувалися кількісні методи. Для отримання кількісних даних, їх обробки для дослідження нами було використане середовище Voyant Tools (<https://voyant-tools.org/>).

Для проведення контентного аналізу ми скористалися варіантом змістового контент-аналізу, запропонованого Вернером Фрю (див. [3; 4]). У нашому дослідженні «зоною релевантності» контент-аналізу є ступінь представленості певних категорій у текстах місій університетів. На етапі квантифікації тексту як одиниці контенту використовувалися тексти місій, а як одиниця рахунку – результати кількісної та якісної експертної оцінки текстового матеріалу.

З метою проведення контент-аналізу проголошених місій українських університетів із загальної кількості – 1043 – закладів вищої освіти (далі у тексті – ЗВО), що зазначені у «Реєстрі суб'єктів освітньої діяльності: заклади вищої освіти», ми вибрали ті

університети України, які потрапили до рейтингу «ТОП-200 Україна 2021». Таких університетів у 2021 році виявилось 201, тому що 2 університети мали ідентичні показники і, отже, розділили одне (147 місце) рейтингу. Тексти місій були взяті з офіційних веб-сайтів університетів. За результатами аналізу вмісту сторінок з вибірки ми виключили 12 університетів, місії яких у загальнодоступних джерелах нам виявити не вдалося. 189 ЗВО, що залишилися, ми розділили на 9 груп за їх профільністю (таблиця 1 з [9]). Вибрані нами профілі університетів відповідають статті 28 Закону України «Про вищу освіту». Аналіз усередині профілів ЗВО є важливим, оскільки його результати можуть значно відрізнитися (нами це було визначено в [2]).

На першому етапі контент-аналізу нами використовувалися кількісні методи аналізу для визначення слів, що найчастіше зустрічаються в текстах місій, а також для визначення взаємозв'язку профільності університетів і кількості символів і слів у текстах місій.

Як категорії другого етапу контент-аналізу текстів місій вибірки університетів України нами були визначені такі:

А) навчальна складова місії з 4-ма полями для аналізу тексту заяв – 1) заява про формування нових способів мислення; 2) декларування створення академічного середовища; 3) декларування створення освітніх послуг; 4) заява про намір просування знань у всьому світі;

Б) дослідницька складова місії з 4-ма полями для аналізу тексту заяв – 1) заява про глобальний контекст досліджень; 2) заява про виконання передових досліджень; 3) заява про інноваційний характер результатів досліджень; 4) декларування міжнародної відкритості та співробітництва у дослідженнях;

В) складова служіння суспільству з трьома полями для аналізу тексту заяв – 1) декларування передачі технологій та інновацій у неакадемічне середовище; 2) заява про підтримку безперервної освіти протягом усього життя; 3) декларування намірів щодо соціальної залучення до взаємодії з громадянськими, культурними, промисловими та діловими спільнотами.

Для експертної оцінки

наявності/відсутності заяв у місіях університетів у розрізі позначених полів категорій першого етапу контент-аналізу використовувалася 2-бальна система (наявність – 1, відсутність – 0).

Як категорії третього етапу контент-аналізу текстів місій було визначено такі показники (властивості місій): А) відкритість (доступність) тексту місії; Б) відповідність історичного контексту заяви місії та «віку» університету; В) відповідність заяв про міжнародну відкритість та наявність перекладів тексту місії; Г) визначення у тексті місії зацікавлених у співпраці сторін; Д) позначення у тексті місії унікальності навчального закладу; Е) визначення у тексті місії сильних сторін навчального закладу; Ж) лапідарність тексту місії; З) «довговічність» заяви місії. Для експертної оцінки наявності/відсутності властивостей заяв у місіях університетів у розрізі позначених полів категорій другого етапу контент-аналізу використовувалася також 2-бальна система (наявність – 1, відсутність – 0).

За результатами вивчення процесу відображення особливостей освітнього, наукового й соціального середовища в текстах місій закладів вищої освіти України зроблені такі висновки.

1. Аналіз психолого-педагогічних, філософських та соціологічних джерел показав, що початок дискусії щодо ідеї університету було покладено Джоном Ньюменом у циклі лекцій «Ідея університету» (1873). Однак концептуальні основи університетської освіти ще раніше були сформульовані та втілені в життя німецьким енциклопедистом та реформатором Берлінського університету В. Гумбольдтом. Серед авторів публікацій на цю тему, які вже були згадані вище текстом, – Ф. Ніцше, Х. Ортега-і-Гассет, К. Ясперс, А.М. Уайтхед та багато інших. В результаті сформувалися суттєві відмінності у змісті ідеї університету. Втім, саме позиція, яка трактує триєдину – наукову, навчальну та культурну – місію університету – найближча до української ідеї університету. В.І. Вернадський вважав університет центром національної культури – «вища школа має культивувати і найвище знання». Друга характеристика закладу вищої освіти – свобода

думки – «відсутність зовнішнього примусу». Третя виділена ним характеристика – громадське служіння. У результаті ми прийшли до висновку, що освіта в сучасному світі, завдяки її функціям, відіграє інтегративну роль при формуванні соціальних спільнот, соціальних інститутів суспільства, формуванні взаємодіючої особистості. Освіта є єдиним інститутом суспільства, на якому будується найбільша кількість взаємовідносин між соціальними спільнотами та їх представниками.

2. Проведений аналіз еволюції поняття місії університету в часовому та територіальному вимірі продемонстрував неоднорідність та неодночасність її формування у різних освітніх суб'єктів як у часовому проміжку, так і територіально. В результаті вдалося з'ясувати, що місія закладу вищої освіти бачиться як його основна мета, сенс чи причина, через яку інститут існує у суспільстві. Вона вважається основою, завдяки якій установа будує своє бачення чи свої стратегічні плани. Заяви про місію виходять із внутрішнього середовища установи та враховують потреби внутрішніх зацікавлених сторін, а також визначаються тиском навколишнього середовища та проблемами, які можуть вплинути на організацію. Створюючи ефективне формулювання місії, університет передає свою філософію та основні компетенції, які допоможуть досягти його мети.

3. Проведення порівняльного контентного аналізу змісту місії українських закладів вищої освіти в загальному дало можливість переконатися, що традиції, закладені у сфері українського навчання та дослідження, сприйняті педагогічними університетами: навчальна та дослідницька функції заявляються як пріоритетні в місіях більшості закладів вищої освіти. Крім того, у місіях університетів знайшли відображення традиційні для університетів теми: служіння суспільству, орієнтація на передові дослідження, формування академічного середовища, виховання у студентів здатності мислити, підготовка кваліфікованих кадрів.

4. Проведення першого етапу контент-аналізу показало відсутність зв'язку між профільністю університетів вибірки та

обсягом тексту місії (як кількості слів, так і кількості знаків). Разом з тим, при вивченні статистики використання термінології стали виявлятися відмінності, пов'язані з профільністю ЗВО, отже, з областями знань, що мають різний термінологічний глосарій викладу.

5. Проведення другого етапу контент-аналізу показало, що у змістовній частині заяв більш ніж 70 % університетів присутні навчальна, дослідницька та соціальна складові текстів місії. Найбільш активно університети робили заяви про створення освітнього середовища (навчальна складова місії), проведення передових досліджень (дослідна складова), а також декларували наміри про соціальну залученість у взаємодію з громадянськими, культурними, промисловими та діловими спільнотами (соціальна складова). Причому декларування створення академічного середовища змістовно варіювалося від визначення конкретних факторів та умов до заяв загального змісту. Заяви щодо виконання передових досліджень також відрізнялися за змістом: від визначення конкретних галузей науки до узагальнених формулювань. Декларації про наміри залучення ЗВО до соціальної взаємодії з різними спільнотами враховували регіональний, загальнодержавний чи міжнародний рівні впливу.

6. Найбільш популярною складовою текстів місії університетів у порівнянні з іншими складовими є соціальна. При цьому слід зазначити, що університети явно недостатньо приділили увагу в текстах декларуванню передачі неакадемічному середовищу технологій та інновацій (де такі знання можуть принести соціальні та комерційні вигоди на місцевому, регіональному, національному чи глобальному рівнях), а також заявам намірів підтримки безперервної освіти протягом всього життя (що могло стати мотивацією до поліпшення знань, навичок та компетенцій і, отже, призвести до підвищення якості особистої, громадянської та соціальної зайнятості). Цей недолік може спричинити зниження маркетингового потенціалу місії.

7. Більшість університетів України недооцінили можливість збільшення маркетингового ефекту тексту місії за наявності в них заяв про використання в навчальному процесі сучасних освітніх сервісів та намірів просування знань по всьому світу, а також за наявності декларацій про наміри надати (або констатувати) глобальний характер науковим дослідженням та заяв про організацію наукової діяльності у рамках міжнародного співробітництва.

8. Результати третього етапу контент-аналіз показали, що заяви університетів вибірки переважно визначають сильні сторони навчальних закладів, можуть служити установам протягом ряду років і на тлі загальних статистичних показників більшість із них задовольняє умову лапідарності. При цьому чітко проглядають резерви для покращення текстів місій університетів України. І насамперед у контексті обґрунтованого звернення до історичного досвіду заснування та інтернаціоналізації заяви, а також у плані доступності тексту, визначення у ньому зацікавлених у співпраці сторін та позначення унікальності ЗВО. Втім, враховуючи результати вивчення наявності зв'язку заяв у місіях у розрізах другого та третього етапів контент-аналізу, слід визнати, що при написанні текстів місій більшістю університетів не останню роль відігравали співробітництво у новому для них заході та взаємне консультування.

РОЗДІЛ 3.

ДЕСКРИПТИВНА СТАТИСТИКА. ВИБІРКОВИЙ МЕТОД

3.1. Теоретичні відомості: описова статистика, вибірки, групування

Практично кожен історик зустрічається у своїй роботі з необхідністю обробки та подальшої презентації статистичних даних. Причому це стосується не лише історичних досліджень, а й найрізноманітніших галузей науки – від технічних та медичних до соціологічних та культурологічних. Зумовлена така поширеність тим, що статистичні методи допомагають отримати й обґрунтувати певні судження про об'єкти, суб'єкти, групи людей та інші структури, що мають певну внутрішню неоднорідність.

Ті дослідники, хто у своїй роботі стикається з обробкою та описом даних, мають чітко уявляти, в якій формі їх коректно відображати та обробляти для подальшого використання при формулюванні статистичних висновків. Тому слід розуміти, що таке описова статистика. Вона ще зветься дескриптивною і займається аналізом та обробкою емпіричних даних з проведенням необхідної систематизації. Отже, *описова статистика* – це стисла й концентрована характеристика явища, що вивчається, яка представлена у вигляді графіків, таблиць, схем і числових виразів.

До описової статистики в якості основних показників входять змінні, які можна не тільки вимірювати, а й змінювати в ході

певних маніпуляцій; екстремуми, або так звані максимуми та мінімуми значень самої змінної; варіаційні та часові ряди, що містять усі кількісні ознаки, що є у кожній одиниці статистичного спостереження; середні значення, які можуть бути представлені середнім гармонійним, геометричним, арифметичним, зваженим та квадратичним – всі вони потрібні для того, щоб охарактеризувати центр розподілу; мода – значення, яке найчастіше зустрічається у вибірці; медіана – середні значення упорядкованих вибірок; дисперсія – дозволяє оцінити відхилення у певному числі спостережень.

Крім цього, для реалізації методів описової статистики використовують такі показники, як квартилі, асиметрію, статистичні моменти, ексцеси, гіпотези, значимості. Кожен з них відіграє істотну роль коректного відображення одержуваних даних.

Отже, описова, або *дескриптивна статистика* – це галузь статистики, яка вивчає методи збору, організації, опису та візуалізації даних з метою розуміння їх основних характеристик. Основна мета дескриптивної статистики полягає в тому, щоб узагальнити та інтерпретувати інформацію, що міститься у наборі даних. Основні поняття та методи, які використовуються в дескриптивній статистиці, включають:

1. Центральні тенденції – міри центральної тенденції, такі як середнє арифметичне, медіана та мода, використовуються для визначення «середнього» значення набору даних.

2. Розмах – це різниця між найбільшим і найменшим значеннями у наборі даних, що вказує на ступінь варіабельності в даних.

3. Розподіл частот, який показує, як часто з'являється кожне значення в наборі даних.

4. Міри розсіювання, такі як дисперсія та стандартне відхилення, використовуються для вимірювання ступеня варіації в наборі даних.

5. Гістограми та графіки – дескриптивна статистика часто використовує графічні засоби для візуалізації розподілу та

взаємозв'язків між даними, такі як гістограми, діаграми розсіювання та коробкові діаграми.

6. Квартилі, які допомагають розуміти розподіл даних, вказуючи значення, які ділять набір даних на рівні частини (наприклад, медіана, квартилі, децилі).

Сукупність вище наведених показників допомагає при візуальному поданні даних здійснити фіксацію їх щодо осей, надавши їм вагу у числовому вираженні; проілюструвати розмах даних; показати асиметричність розподілу відносно його центру; вивести закон розподілу даних за допомогою гістограми, таблиці частот чи функції. Загалом, дескриптивна (або описова) статистика дозволяє дослідникам отримати перші уявлення про характеристики набору даних, розкриваючи їхню структуру та основні властивості. Це важливий етап в аналізі даних перед застосуванням більш складних статистичних методів.

Найчастіше історики мають у своєму розпорядженні великий масив джерел та даних, який називається *генеральною сукупністю* даних. Але досить часто вони не в змозі повністю обробити великі обсяги даних. Це стосується, насамперед, досліджень з Нової та Новітньої історії. З іншого боку, чим глибше доводиться занурюватися в глибину століть, то меншою кількістю інформації можна оперувати. В обох випадках доцільно використовувати так званий вибірковий метод, суть якого полягає в заміні суцільного обстеження масових однорідних об'єктів частковим їх дослідженням. При цьому з генеральної сукупності виділяється частина елементів, що називається *вибіркою*, і результати обробки вибірових даних у результаті узагальнюються на всю сукупність. Основою для характеристики всієї сукупності може бути лише репрезентативна вибірка, яка правильно відбиває властивості генеральної сукупності. Це досягається методом випадкового відбору елементів генеральної сукупності, при якому всі її елементи мають рівні шанси потрапляння до вибірки.

Застосування даного методу однаково підходить і для вивчення різних явищ і процесів сучасності, і для обробки даних проведених

раніше вибіркового статистичного дослідження, наприклад, переписи. Крім того, вибіркового метод також знаходить застосування при обробці даних природних вибірок, від яких залишилися лише дані, що збереглися фрагментарно. Так, досить часто до таких частково збережених даних належать актові матеріали, документи поточного діловодства та звітності.

Залежно від цього, яким чином здійснюється відбір елементів сукупності до вибірки, розрізняють кілька видів вибіркового обстеження, у яких відбір може бути випадковим, механічним, типовим і серійним.

Випадковим називається відбір, коли всі елементи генеральної сукупності мають рівну можливість бути відібраними, наприклад, з допомогою жереба чи таблиці випадкових чисел.

Спосіб жеребкування застосовується в тому випадку, якщо кількість елементів всієї сукупності невелика. При великому обсязі даних здійснення випадкового відбору за допомогою жеребкування стає складним. Найбільш придатним у разі великого обсягу оброблюваних даних є метод використання таблиці випадкових чисел (див. [1]).

Спосіб відбору за допомогою таблиці випадкових чисел можна розглянути на прикладі. Припустимо, що сукупність складається з 900 елементів, а намічений обсяг вибірки дорівнює 20 одиницям. У такому випадку з таблиці випадкових чисел слід відбирати числа, що не перевищують 900, доки не будуть набрані необхідні 20 чисел. Виписані числа слід вважати порядковими номерами елементів генеральної сукупності, що потрапили до вибірки.

Для великих сукупностей краще застосувати *механічний відбір*. Так, при формуванні 10% вибірки з кожних десяти елементів вибирається тільки один, а вся сукупність умовно розбивається на рівні частини по 10 елементів. Далі з першої десятки навмання вибирається якийсь елемент (наприклад, за допомогою жеребкування). Інші елементи вибірки визначаються зазначеною пропорцією відбору n -им номером першого відібраного елемента.

Ще одним видом спрямованого відбору є *типовий відбір*, коли сукупність розбивається на групи, однорідні у якісному відношенні. Тільки після цього всередині кожної групи проводиться випадковий відбір. Хоча це й складніший метод, він дає більш точні результати.

Серійний відбір є видом випадкового або механічного відбору, що здійснюється для укрупнених елементів вихідної сукупності, яка в ході аналізу розбивається на групи (серії).

Дескриптивні методи досліджень використовуються для збору, опису і висновків про властивості явища або популяції без розробки статистичних гіпотез. Ці методи дозволяють відобразити та описати досліджувану групу або явище, розкриваючи його основні характеристики. Описові методи досліджень включають:

1. Опитування – це метод збору даних, при якому дослідник задає питання учасникам дослідження для отримання інформації про їхні думки, переконання, досвід або характеристики.

2. Спостереження – включає систематичний аналіз та фіксацію подій, явищ і взаємодій в природному оточенні без активного втручання дослідника.

3. Архівні дослідження – дослідники можуть аналізувати архівні документи, дані або ресурси, щоб отримати інформацію про минулі події або тренди.

4. Крос-секційні дослідження – вивчають характеристики певної групи людей, об'єктів або явищ у певний момент часу.

5. Кореляційні дослідження – цей метод дослідження вивчає зв'язки між різними змінними без втручання дослідника.

6. Дослідження ретроспективного аналізу – дослідники вивчають минулі події, щоб з'ясувати їхній вплив на сьогоденні стан речей.

Отже, описові методи досліджень допомагають отримати загальне уявлення про об'єкт дослідження, відповісти на питання "що?", "де?", "коли?" і "як?" і надати основу для подальшого аналізу та висновків.

Кількісні методи досліджень використовуються для збору та аналізу кількісних даних, які можуть бути виміряні та оброблені за допомогою статистичних методів. Ці методи спрямовані на виявлення зв'язків, встановлення закономірностей та перевірку гіпотез. Деякі типи кількісних методів досліджень включають:

1. Експериментальні дослідження – дозволяють дослідникам встановлювати причинно-наслідкові зв'язки, порівнюючи вплив різних умов чи обставин на досліджувану змінну.

2. Опитування з анкетуванням – є одним із найпоширеніших методів збору даних у кількісних дослідженнях. Вони використовуються для збору стандартизованої інформації від великої кількості учасників.

3. Секційні дослідження – тут вивчаються характеристики одного і того ж об'єкту або їхньої групи протягом певного періоду часу, щоб виявити зміни та тенденції в часі.

4. Кореляційні дослідження – вивчають взаємозв'язки між двома або більше змінними, використовуючи статистичні методи для визначення сили та напрямку цих зв'язків.

5. Експертні оцінки – використовуються для отримання думок та оцінок від експертів у певній області знань.

6. Соціометрія – вивчає взаємодії та структури в групах, використовуючи кількісні методи для аналізу соціальних мереж та взаємодій.

Ці методи дозволяють отримувати кількісні дані, які можуть бути проаналізовані за допомогою статистичних методів для отримання об'єктивних результатів та висновків.

У кількісних дослідженнях використовуються різні *шкали вимірювань*, які дозволяють оцінювати різні аспекти досліджуваних явищ.

1. Номінальна шкала – використовується для класифікації об'єктів чи учасників на категорії без встановлення будь-якого порядку чи ієрархії. Наприклад, шкала «стать» (чоловік/жінка) або «кольори світлофора» (червоний/жовтий/зелений).

2. Порядкова шкала – дозволяє встановлювати порядок чи ранжування між категоріями, але відстань між категоріями не вимірюється. Наприклад, шкала «згоджуюсь/не згоджуюсь/не впевнений» або «рівні ступені задоволення» (низький/середній/високий).

3. Інтервальна шкала – має фіксовані відстані між значеннями, але не має абсолютного нуля. Наприклад, температура в Цельсіях: 20°C, 30°C, 40°C тощо. Важливо зауважити, що на інтервальній шкалі можна виконувати операції додавання та віднімання.

4. Відсоткова шкала – це варіант інтервальної шкали, але з абсолютним нулем. Наприклад, вік у роках: 0 років означає відсутність віку. На відсотковій шкалі можна виконувати всі арифметичні операції.

5. Абсолютна шкала має фіксовані відстані між значеннями та абсолютний нуль. Наприклад, вага у кілограмах: 0 кг означає відсутність ваги, і на цій шкалі можна виконувати всі арифметичні операції.

Такі шкали використовуються для вимірювання різних аспектів досліджуваних явищ і дозволяють дослідникам отримати різні рівні даних для аналізу та інтерпретації.

Кожен факт, кожна історична особа характеризуються унікальним набором ознак. Втім у процесі вивчення та аналізу виявляється спільність у ознакових показниках. Як правило, значення тих самих ознак різних явищ або повторюються, або слабо розрізняються. Ця спільність дозволяє низку явищ об'єднувати в одну групу. Яку б сукупність об'єктів ми не розглядали, її можна розбити на групи за подібністю ознак. Так, при всій унікальності кожного людського обличчя в сукупності можна визначити й назвати його типи (монголоїдний, привітний, овальний і т.п.)

Вивчаючи явища минулого за первинними статистичними даними, історик стикається з неупорядкованою послідовністю чисел, показників, що характеризують той чи інший аспект явища чи процесу. Одним з найбільш поширених прийомів подання

сукупності розрізнених даних у зручній для сприйняття формі є їхнє *групування*. Процес групування є основним початковим етапом обробки даних джерела, фундаментом більшості інших прийомів математико-статистичного аналізу.

Метод групування полягає у розбиття вихідної сукупності даних на групи, кожна з яких об'єднана загальними показниками. Відмінності між елементами однієї групи мають бути меншими, ніж відмінності між елементами різних груп.

Згруповані дані подаються у вигляді таблиць чи графіків. Це дозволяє охарактеризувати як загалом досліджувану сукупність, так і її частини; виявити та зафіксувати зв'язки між ознаками; забезпечити наочність та компактність презентації матеріалу.

Найвний у розпорядженні дослідника набір даних (як правило, чисел) називається статистичною сукупністю. Кількісні показники, які характеризують ознаку, що розглядається, і приймають різні значення називають варіантами або змінними. Відомості джерела, систематизовані у зростаючому чи спадному порядку та оформлені як таблиці називаються *варіаційним рядом*.

Варіаційні ряди, або ряди варіацій, використовуються в статистиці для організації та представлення даних про розподіл значень змінної. Основною метою варіаційних рядів є розкриття розмаху значень змінної та їх частот.

Існує два основних типи варіаційних рядів:

1. Простий варіаційний ряд – в простому варіаційному ряді значення змінної впорядковані за зростанням або спаданням, і порядок відображає розподіл даних. Кожне значення супроводжується відповідною частотою – кількістю разів, які це значення з'являється у наборі даних.

2. Групований варіаційний ряд – у групованому варіаційному ряді значення змінної розділені на групи (інтервали) з однаковою шириною, або інтервалами, що відрізняються. Кожен інтервал має відповідну частоту – кількість спостережень, які потрапляють у цей інтервал. Групований варіаційний ряд дозволяє візуалізувати розподіл даних у вигляді гістограми або полігону частот.

Варіаційні ряди допомагають визначити розмах і розподіл значень змінної в наборі даних. Вони є важливим інструментом для визначення центральних та розмахових мір, вивчення структури розподілу даних та виявлення відмінностей між різними групами або категоріями.

Ознаки, покладені в основі складання таблиці, можуть бути дискретними, тобто такими, що приймають лише цілі значення, і неперервними, якщо окремі значення можуть відрізнятися один від одного на скільки завгодно малу величину. Прикладом дискретної ознаки може бути кількість дітей у сім'ї, а безперервної – стаж роботи.

У практиці історичних досліджень найчастіше використовують таблиці з інтервальною розбивкою ознаки. Для того, щоб не втратити інформацію і в той же час скласти компактну таблицю, використовують інтервальні ряди. Тут перед дослідником постає проблема визначення меж інтервалів. Необхідно знайти оптимальну кількість груп, кількість інтервалів ознаки та встановити розмір інтервалів.

Розв'язання цього завдання залежить від ступеня однорідності аналізованої сукупності. У разі, якщо сукупність однорідна, рекомендується брати рівні інтервали. Слід пам'ятати, що при описі тенденції у розподілі змінних ознаки інтервали краще укрупнити. У разі, коли значення мають конкретні дані щодо кожної групи, інтервали має сенс зробити невеликими. Таким чином, вибір інтервалів залежить від властивостей досліджуваного процесу чи явища та від мети роботи, причому питання це вирішується змістовним, якісним аналізом та залежить від професійних навичок історика.

Проте, є кілька формальних способів визначення оптимальної величини інтервалу, тобто. такого його значення, у якому проглядалася б специфіка явища й водночас групування б не було громіздким. Найбільш проста у вживанні формула, запропонована Г. Стерджессом:

$$K = \frac{X_{max} - X_{min}}{1 + 3,2 \ln n}$$

де K – величина інтервалу; X_{max} – найбільше значення ознаки; X_{min} – найменше значення ознаки; n – число елементів сукупності.

Користуючись даними, групування яких зроблено статистиками-професіоналами, ми повинні знати, що розбиття групувальної ознаки виконується таким чином, щоб розподіл частот у кожній групі був приблизно рівним. У розв'язанні багатьох завдань історика краще користуватися первинними, незгрупованими матеріалами, здійснюючи групування та перегрупування даних самостійно, відповідно до мети свого дослідження. Границі інтервалів для дискретних ознак встановлюються без співпадіння крайніх показників суміжних інтервалів. Це пояснюється властивостями самої ознаки, що приймає лише цілі значення. Отже, при співпадінні крайніх значень сусідніх інтервалів один і той самий показник може бути зафіксований у двох інтервальних групах. Для дробових (безперервних) ознак застосовується протилежне правило – обов'язковий збіг суміжних границь інтервалів.

У разі неоднорідної сукупності об'єктів вивчення в основу побудови групування закладається якісний критерій, покликаний виявити однорідні типи. Цей принцип спрямований на те, щоб визначити межі інтервалів там, де кількісна зміна ознаки призводить до появи нової якості. У разі потреби сукупність розбивається на однорідні групи, усередині кожної будується своя шкала інтервалів.

Формальні математико-статистичні методи неспроможні надати істотну допомогу історика у виборі принципу групування. Це прерогатива якісного аналізу. Метод групування дозволяє складне явище уявити через ряд більш простих, що допомагає прийти до аналізу всієї системи в цілому. Метод сприяє оцінці інформаційного потенціалу джерела. Методом групування характеризуються типи явищ у їхніх взаємних відносинах, а також

розкривається причинна залежність між окремими факторами й загальною тенденцією розвитку процесу.

У науці розрізняють 3 основні *види групувань*.

Типологічні – розчленовують якісно-різнорідну сукупність на однорідні групи, типи. В основу групування закладається якісна ознака.

Структурні – представляють якісно однорідну сукупність у вигляді кількісних груп. В основу цих згрупувань закладається кількісна ознака.

Одним з найважливіших моментів використання групувального методу й визначення виду групування є вибір групувальної ознаки. Від прийому групування, від базової ознаки багато в чому залежать висновки, які можна отримати на основі тих самих матеріалів, які аналізувалися одним і тим же методом групування. Головна вимога до групувальної ознаки – достовірність відображення структури явища, що вивчається в залежності від часу і конкретно-історичних умов. Виконання цієї вимоги залежить від професійних навичок історика, тобто від якісного боку дослідження.

Розподіл згрупувань на типологічні та структурні відносно й залежить від характеру завдань, що стоять перед дослідником. Наприклад, якщо задати межі землекористування, що відповідають певним соціальним групам селянства, то можна вивчити структуру селянських господарств за розмірами землекористування та типи господарств щодо розмірів землекористування. Таким чином, відмінність полягає у підрозділах групувальної ознаки.

Третій вид – *аналітичні групування*. Вони дозволяють встановити та на певному рівні вивчити взаємозв'язок між ознаками. У статистичній літературі такі групування ще називають факторними, причому одна з групувальних ознак розглядається як результат, а інша – як фактор. Вивчення та інтерпретація даних аналітичного угруповання повинні починатися з попереднього з'ясування принципової можливості існування зв'язку між ознаками.

Особливе місце серед групувань займають часові (динамічні) ряди, що відображають зміну явища в часі. Для історика це має особливе значення, так як найчастіше ми вивчаємо процес, що має певну протяжність у часі.

Часові ряди – це послідовності даних, що вимірюються або збираються в різні моменти часу з фіксованими інтервалами. Ці ряди можуть бути зібрані в різній частоті, наприклад, щоденно, щомісяця, щокварталу тощо. Основними характеристиками часових рядів є часові мітки (або індекси), які вказують час збирання даних, та значення змінної, яка вимірюється у кожній точці часу. Часові ряди використовуються для аналізу та передбачення змін змінних в часі. Крім історичних досліджень вони широко застосовуються в таких областях, як економіка, фінанси, метеорологія, соціологія, наука про дані та інші. Розрізняють такі види часових рядів:

1. Трендові ряди – ці ряди показують загальну тенденцію зміни значення змінної протягом тривалого періоду часу. Наприклад, зростаюча кількість продажів товарів протягом кількох років.

2. Сезонні ряди – ці ряди відображають періодичні зміни в значеннях змінної, які повторюються в одному або декількох сезонах або періодах. Наприклад, щорічні коливання зборів сільськогосподарської продукції відповідно до сезону року.

3. Циклічні ряди – ці ряди відображають довгострокові циклічні коливання в значеннях змінної, які можуть бути пов'язані з економічними циклами або іншими фундаментальними факторами.

4. Випадкові ряди – такі ряди представляють випадкові коливання в значеннях змінної без виражених трендів, сезонності або циклів.

Часові ряди дозволяють аналізувати та передбачати поведінку даних в часі, що є важливим для багатьох аспектів діяльності та прийняття рішень. Включення до групування хронологічного фактору зумовлює специфічні вимоги до його побудови та методів аналізу. Час у часових рядах може бути задано у вигляді

конкретних дат (моментів часу) або у вигляді проміжків – років, місяців, доби. Показники часового ряду називаються рівнями.

Дослідження динамічних рядів починається з аналізу вимог, що пред'являються до їхньої побудови.

1. Явища щодо кожної динамічної групи мають бути однорідними, тобто в один часовий проміжок повинні включатися одні й ті ж явища.

2. Територія, до якої належать показники, має бути незмінною.

3. Методологія обліку показників має бути єдиною. Рівні динамічного ряду можуть бути задані або абсолютними, або відносними, або середніми величинами (відповідно підрозділяються ряди динаміки).

Історик, який має справу з відносними чи середніми величинами, повинен перевірити історію їх виникнення, тобто з'ясувати – у яких одиницях підраховувалися наявні показники.

4. Часовий показник, покладений в основу часового ряду у разі його інтервальної розбивки, повинен мати порівнянні часові проміжки.

Показники динамічного ряду вважаються порівнянними при виконанні всіх чотирьох умов його побудови.

При вимірюванні кількісних ознак важливо дотримуватися кількох ключових принципів та звертати увагу на певні аспекти.

1. Точність вимірювань – важливо, щоб вимірювання були якомога точнішими та надійними. Це означає використання точних та каліброваних інструментів, правильне виконання процедур вимірювання та контроль за потенційними джерелами помилок.

2. Репрезентативність вибірки – важливо, щоб вибірка, яку ви використовуєте для вимірювань, була репрезентативною для популяції, яку ви досліджуєте. Це допоможе зробити узагальнення результатів на всю популяцію.

3. Співмірність міри – при вимірюванні різних об'єктів або явищ важливо використовувати однакові міри для всіх об'єктів.

Наприклад, якщо вимірюється вага одного предмета в грамах, то всі інші предмети також мають бути вимірювані в грамах.

4. Стандартизація процедур – важливо мати стандартизовані процедури вимірювання, які застосовуються до всіх об'єктів. Це допомагає уникнути систематичних помилок та забезпечує однакові умови для всіх вимірювань.

5. Обробка даних – після вимірювань важливо правильно обробляти та аналізувати отримані дані. Це може включати розрахунок середніх значень, стандартного відхилення, кореляційних коефіцієнтів та інших статистичних показників.

6. Перевірка вимірювань на відхилення та аномалії – важливо виявляти та враховувати будь-які відхилення чи аномалії в даних. Це може вказувати на можливі помилки в процесі вимірювання або на особливості досліджуваного явища.

Загальний принцип полягає в тому, щоб забезпечити точність, достовірність та репрезентативність вимірювань для досліджуваної проблеми.

Варіаційні та часові ряди наочно представляються графічно. Графічне представлення варіаційних рядів є важливим інструментом для візуалізації розподілу даних та виявлення основних характеристик. Деякі з основних способів графічного представлення варіаційних рядів включають:

1. Гістограму, яка представляє розподіл частот (або відсотків) значень змінної відносно їхніх категорій або інтервалів. Кожний стовпець гістограми відповідає певному інтервалу значень, а висота стовпця відображає частоту (або відсоток) значень в цьому інтервалі.

2. Полігон частот – це графічне представлення розподілу частот або відсотків значень змінної, де кожна точка показується на графіку, і лінія з'єднує ці точки.

3. Кумулятивну гістограму, яка показує накопичену частоту (або відсоток) значень менших за певну величину. Вона створюється шляхом послідовного додавання частот відповідних інтервалів.

4. Діаграму розсіювання, яка використовується для візуалізації взаємозв'язку між двома змінними. Кожна точка на графіку відображає пару значень цих змінних.

5. Коробкову діаграму – ця діаграма показує розподіл значень змінної, а також основні характеристики такі як медіана, квартилі та виявлення викидів.

6. Лінійний графік, який використовується для відображення змін значення змінної відносно часу або іншої незалежної змінної. Кожна точка на графіку показує значення змінної в певний момент часу або відповідно до іншої змінної.

Ці графіки допомагають аналізувати та порівнювати розподіл даних, виявляти взаємозв'язки та тренди, а також робити висновки про характеристики досліджуваної змінної.

Графічне представлення часових рядів є важливим інструментом для візуалізації динаміки даних в часі та виявлення закономірностей та трендів.

1. Лінійний графік – це найбільш поширений спосіб візуалізації часових рядів. На графіку відображаються значення змінної в залежності від часу, де вісь X представляє часові моменти, а вісь Y – значення змінної.

2. Графік зі згладжуванням (середній графік) – цей графік показує середнє значення змінної протягом певного періоду часу, що допомагає виявити основні тенденції та відфільтрувати випадкові коливання.

3 Кумулятивний графік – цей графік відображає накопичену суму значень змінної протягом часу. Він корисний для візуалізації накопичених змін у даних.

4. Графік в розмаху – цей графік показує основні статистичні показники (наприклад, медіану, квартилі, мінімальне та максимальне значення) змінної протягом певного періоду.

5. Діаграма розсіювання, яка використовується для візуалізації взаємозв'язку між двома змінними в часі. Кожна точка на графіку відповідає певному часовому моменту та відповідному значенню змінної.

6. Сезонні графіки – ці графіки використовуються для виявлення та візуалізації сезонних закономірностей у часових рядах. Вони допомагають ідентифікувати регулярні коливання в даних, пов'язані з певними періодами часу, такими як місяці, квартали або роки.

Ці способи графічного представлення допомагають зрозуміти динаміку часових рядів, виявити тенденції та сезонність, а також визначити взаємозв'язки між різними змінними.

Отже, метод групування є необхідним початковим етапом аналізу масових джерел. Він розв'язує завдання встановлення статистичних зв'язків і закономірностей, виявлення структури сукупності, що вивчається, визначення та описи її типів. З його допомогою складне явище чи подія постає у вигляді кількох простіших. Метод тісно пов'язаний з табличною та графічною формою подання даних. Таблиці та їхня візуалізація наводяться з метою подальшого аналізу інформаційної бази, для ілюстрації отриманих висновків, для чіткого опису суті проблеми.

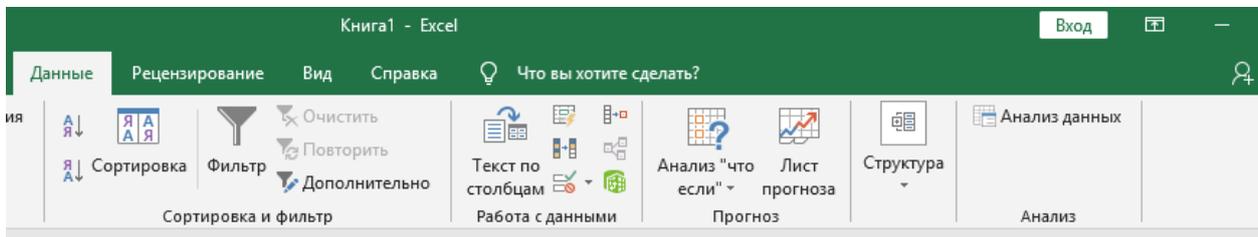
3.2. Застосування інструменту EXCEL Описова статистика даних вибірки

Розглянемо інструмент *Описова статистика*, що входить до надбудови *Пакет Аналізу*. Для цього розрахуємо показники вибірки: середнє, медіана, мода, дисперсія, стандартне відхилення та ін.

Завдання описової статистики (descriptive statistics) полягає в тому, щоб з використанням математичних інструментів звести сотні значень вибірки до кількох підсумкових показників, які дають уявлення про вибірку. В якості таких статистичних показників використовуються: середнє, медіана, мода, дисперсія, стандартне відхилення та інші.

Опишемо набір числових даних за допомогою певних показників. Нехай дано вибірку: (113, 395, 242, 170, 433, 319, 428, 223, 369, 283, 282, 440, 837, 239, 212, 245, 433, 275, 298, 180, 213, 296, 412, 316, 218, 526, 152, 604, 398, 227).

Для обчислення статистичних показників одновимірних вибірок використовуємо надбудову *Пакет аналізу*. Потім всі показники, розраховані надбудовою, обчислимо за допомогою вбудованих функцій MS EXCEL. Надбудова *Пакет аналізу* (*Analysis ToolPak*) доступна з вкладки *Дані*, група *Аналіз*. Кнопка для виклику діалогового вікна називається *Аналіз даних*.



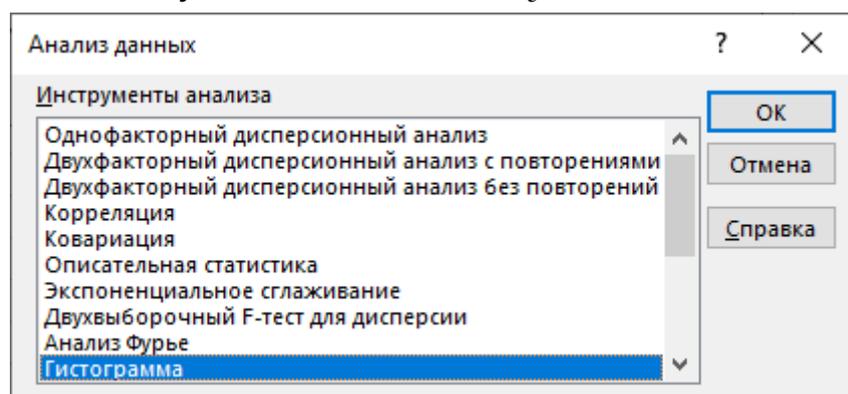
Якщо кнопка не відображається у вказаній групі, необхідно спочатку увімкнути надбудову (нижче дано пояснення для EXCEL 2019):

1. На вкладці *Файл* виберіть пункт *Параметри*, а потім виберіть категорію *Надбудови*.

2. У списку *Керування* (внизу вікна) виберіть *Надбудови Excel* та натисніть кнопку *Перейти*.

3. У вікні *Доступні надбудови* встановіть прапорець *Пакет аналізу* та натисніть кнопку *ОК*.

4. Після натискання кнопки *Аналіз даних* буде виведено діалогове вікно надбудови *Пакет аналізу*.



У діалоговому вікні *Аналіз даних* виберіть інструмент *Описова статистика*. Після натискання кнопки *ОК* буде виведено інше діалогове вікно.

У цьому вікні потрібно вказати:

- *Вхідний інтервал* (Input Range) – це діапазон осередків, у якому міститься масив даних. Якщо в зазначений діапазон входить текстовий заголовок набору даних, потрібно поставити галочку в полі *Мітки в першому рядку* (Labels in first row). У цьому випадку заголовок буде виведено у вихідному інтервалі. Порожні комірки будуть проігноровані, тому нульові значення необхідно обов'язково вказувати в комірках, а не залишати їх порожніми;
- *Вихідний інтервал* (Output Range). Тут вкажіть адресу верхнього лівого осередку діапазону, в який будуть виведені статистичні показники;
- *Підсумкова статистика* (Summary Statistics). Поставте галочку навпроти цього поля – буде виведено основні показники вибірки: середнє, медіана, мода, стандартне відхилення та ін.;

- також можна поставити галочки навпроти полів *Рівень надійності* (Confidence Level for Mean), *K-й найменший* (Kth Largest) та *K-й найбільший* (Kth Smallest).

У результаті для даної вибірки будуть виведені такі статистичні показники:

<i>Стовпчик1</i>	
Середнє	325,9333
Стандартна помилка	27,38474
Медіана	289,5
Мода	433
Стандартне відхилення	149,9924
Дисперсія вибірки	22497,72
Ексцес	3,476698
Асиметричність	1,524488
Інтервал	724
Мінімум	113
Максимум	837
Сума	9778
Рахунок	30
Найбільший(1)	837
Найменший(1)	113
Рівень надійності(95,0%)	56,00808

Пояснимо значення основних показників.

- *Інтервал* (Range) – різниця між максимальним та мінімальним значеннями.
- *Мінімум* (Minimum) – мінімальне значення в діапазоні комірок, зазначеному у Вхідному інтервалі (функція МІН()).

- *Максимум* (Maximum) – максимальне значення (функція МАКС()).
- *Сума* (Sum) – сума всіх значень (функція СУМ()).
- *Рахунок* (Count) – кількість значень у Вхідному інтервалі (порожні комірки ігноруються, функція РАХУНОК()).
- *Найбільший* (Kth Largest) – виводиться найбільший К-й. Наприклад, 1-й найбільший – це максимальне значення (функція НАЙБІЛЬШИЙ()).
- *Найменший* (Kth Smallest) – виводиться К-й найменший. Наприклад, 1-й найменший – це мінімальне значення (функція НАЙМЕНШИЙ()).
- *Середнє вибірки* (mean, average) чи вибіркоче середнє чи середнє вибірки (sample average) є арифметичне середнє всіх значень масиву. У MS EXCEL для обчислення середнього вибірки використовується функція СРЗНАЧ(). Вибіркоче середнє є «хорошою» (незміщеною та ефективною) оцінкою математичного очікування випадкової величини.
- *Медіана вибірки* (Median) – це число, що є серединою множини чисел (у разі вибірки): половина чисел множини більше, ніж медіана, а половина чисел менше, ніж медіана. Для визначення медіани необхідно спочатку відсортувати безліч чисел. Наприклад, медіаною для чисел 2, 3, 3, 4, 5, 7, 10 буде 4. Якщо множина містить парну кількість чисел, то обчислюється середнє для двох чисел, що знаходяться в середині множини. Наприклад, медіаною для чисел 2, 3, 3, 5, 7, 10 буде 4, так як $(3+5)/2$. Якщо є довгий хвіст розподілу, *Медіана* краще, ніж середнє значення, відображає «типове» або «центральне» значення. Для визначення медіани в MS EXCEL існує однойменна функція МЕДІАНА(), англійський варіант – MEDIAN(). *Медіану* також можна обчислити за допомогою формул: = КВАРТИЛЬ. ВКЛ (Вибірка; 2), або = ПРОЦЕНТИЛЬ. ВКЛ (Вибірка; 0,5).

- *Мода вибірки* (Mode) – це значення, що найбільш часто зустрічається (повторюване) у вибірці. Наприклад, у масиві (1; 1; 2; 2; 2; 3; 4; 5) число 2 зустрічається найчастіше – 3 рази. Отже, число 2 – це мода. Для обчислення моди використовують функцію МОДА() , англійський варіант MODE(). Якщо в масиві немає значень, що повторюються, то функція поверне значення помилки #Н/Д. Починаючи з MS EXCEL 2010, замість функції МОДА() рекомендується використовувати функцію МОДА.ОДН(), яка є її повним аналогом. Крім того, в MS EXCEL 2010 з'явилася нова функція МОДА.НСК() , яка повертає кілька значень, що найбільш часто повторюються (якщо кількість їх повторів збігається). Наприклад, у масиві (1; 1; 2; 2; 2; 3; 4; 4; 4; 5) числа 2 і 4 зустрічаються найбільш часто – по 3 рази. Отже, обидва числа є модами. Функції МОДА.ОДН() і МОДА() повернуть значення 2, так як 2 зустрічається першим серед найбільш повторюваних значень. Щоб виправити цю несправедливість і було введено функцію МОДА.НСК(), яка виводить усі моди. Зауважимо, що не дивлячись на те, що мода – це найбільш ймовірне значення випадкової величини (ймовірність вибрати це значення з Генеральної сукупності максимальна), не слід очікувати, що середнє значення обов'язково буде близько до моди.
- *Дисперсія вибірки* або вибіркова дисперсія (sample variance) характеризує розкид значень у масиві, відхилення від середнього (аналітично – це сума квадратів відхилень кожного значення в масиві від середнього, поділена на розмір вибірки мінус 1). У MS EXCEL 2007 і попередніх версіях для обчислення дисперсії вибірки використовується функція ДИСП(). З версії MS EXCEL 2010 рекомендується використовувати її аналог – функцію ДИСП.В(). *Дисперсія вибірки* дорівнює 0 тільки в тому випадку, якщо всі значення дорівнюють між собою і, відповідно, дорівнюють середньому значенню. Чим

більша величина дисперсії, тим більший розкид значень у масиві щодо середнього. Розмірність дисперсії відповідає квадрату одиниці виміру вихідних значень. Наприклад, якщо значення у вибірці є вимірювання ваги деталі (в кг), то розмірність дисперсії буде $кг^2$. Це буває складно інтерпретувати, тому для характеристики розкиду значень частіше використовують величину, що дорівнює квадратному кореню з дисперсії – стандартне відхилення.

- *Стандартне відхилення вибірки (Standard Deviation)*, як і дисперсія – це міра того, наскільки широко розкидані значення у вибірці щодо їхнього середнього. За визначенням, стандартне відхилення дорівнює квадратному кореню з дисперсії. *Стандартне відхилення* не враховує величину значень у вибірці, лише ступінь розсіювання значень навколо їх середнього. У MS EXCEL 2007 і раніше версіях для обчислення *Стандартного відхилення вибірки* використовується функція СТАНДОТКЛОН(). З версії MS EXCEL 2010 рекомендується використовувати аналог СТАНДОТКЛОН.В()).
- У *Пакеті аналізу* під терміном стандартна помилка мається на увазі стандартна помилка середнього (Standard Error of the Mean, SEM). *Стандартна помилка середнього* – це оцінка стандартного відхилення розподілу вибіркового середнього. *Стандартне відхилення розподілу* вибіркового середнього обчислюється за формулою σ/\sqrt{n} , де n – обсяг вибірки, σ – стандартне відхилення вихідного розподілу, з якого взято вибірку. Так як зазвичай стандартне відхилення вихідного розподілу невідоме, то у розрахунках замість σ використовують її оцінку s – стандартне відхилення вибірки. А відповідна величина s/\sqrt{n} має спеціальну назву – *Стандартна помилка середнього*. Саме ця величина обчислюється у *Пакеті аналізу*. У MS

EXCEL стандартну помилку середнього можна також обчислити за формулою =СТАНДОТКЛОН.

- *Асиметричність або коефіцієнт асиметрії (skewness)* характеризує ступінь несиметричності розподілу (щільності розподілу) щодо його середнього. Додатне значення коефіцієнта асиметрії показує, що розмір правого «хвоста» розподілу більший, ніж лівого (щодо середнього). Від'ємна асиметрія, навпаки, вказує на те, що лівий хвіст розподілу більший за правий. Коефіцієнт асиметрії ідеально симетричного розподілу або вибірки дорівнює 0. Функція СКОС(), англійський варіант SKEW(), повертає коефіцієнт асиметрії вибірки, що є оцінкою асиметрії відповідного розподілу.
- *Ексцес* показує відносну вагу «хвостів» розподілу щодо його центральної частини. Зауважимо, що не дивлячись на старання професійних статистиків, у літературі ще трапляється визначення *Ексцесу* як міри «гостроти» (peakedness) або зглаженості розподілу. Але насправді значення *Ексцесу* нічого не говорить про форму піку розподілу. Багато комп'ютерних програм використовують для розрахунків не сам *Ексцес*, а так званий *Kurtosis excess*. Ще більшу плутанину вносить переклад цих термінів українською мовою. Термін *Kurtosis* походить від грецького слова «вигнутий», що «має арку». Так склалося, що українською мовою обидва терміни *Kurtosis* та *Kurtosis excess* перекладаються як *Ексцес* (від англ. excess – «надлишок»). Функція MS EXCEL ЕКСЦЕС() насправді обчислює *Kurtosis excess*. Функція ЕКСЦЕС(), англійський варіант KURT(), обчислює на основі значень вибірки незміщену оцінку ексцесу розподілу випадкової величини. *Ексцес (Kurtosis excess)* для нормального розподілу завжди дорівнює 0.
- *Рівень надійності* означає ймовірність того, що довірчий інтервал містить справжнє значення параметра розподілу,

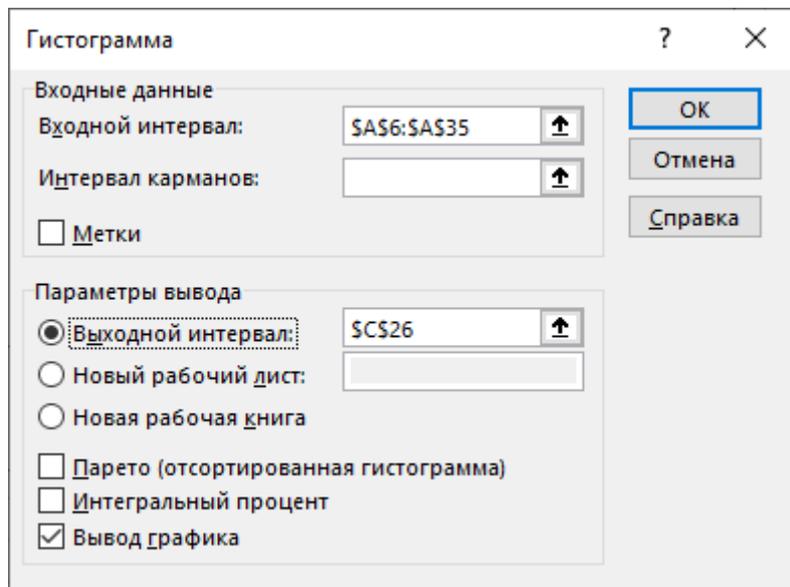
що оцінюється. Замість терміну *Рівень надійності* часто використовується термін *Рівень довіри*. При заданні значення рівня надійності у вікні надбудови *Пакет аналізу*, MS EXCEL обчислить половину ширини довірчого інтервалу для оцінки середнього (дисперсія невідома).

Побудуємо гістограму розподілу. *Гістограма розподілу* – це інструмент, що дозволяє візуально оцінити величину та характер розкиду даних. Створимо гістограму для безперервної випадкової величини за допомогою вбудованих засобів MS EXCEL із надбудови *Пакет аналізу*.

Гістограма (frequency histogram) – це стовпчикова діаграма MS EXCEL, кожен стовпчик якої є інтервал значень (кошик, кишень, class interval, bin, cell), а його висота пропорційна кількості значень у ній (частоті спостережень).

Побудуємо гістограму для набору даних вибірки, де містяться значення безперервної випадкової величини.

Викликавши діалогове вікно надбудови *Пакет аналізу*, виберіть пункт *Гістограма* та натисніть кнопку *ОК*.



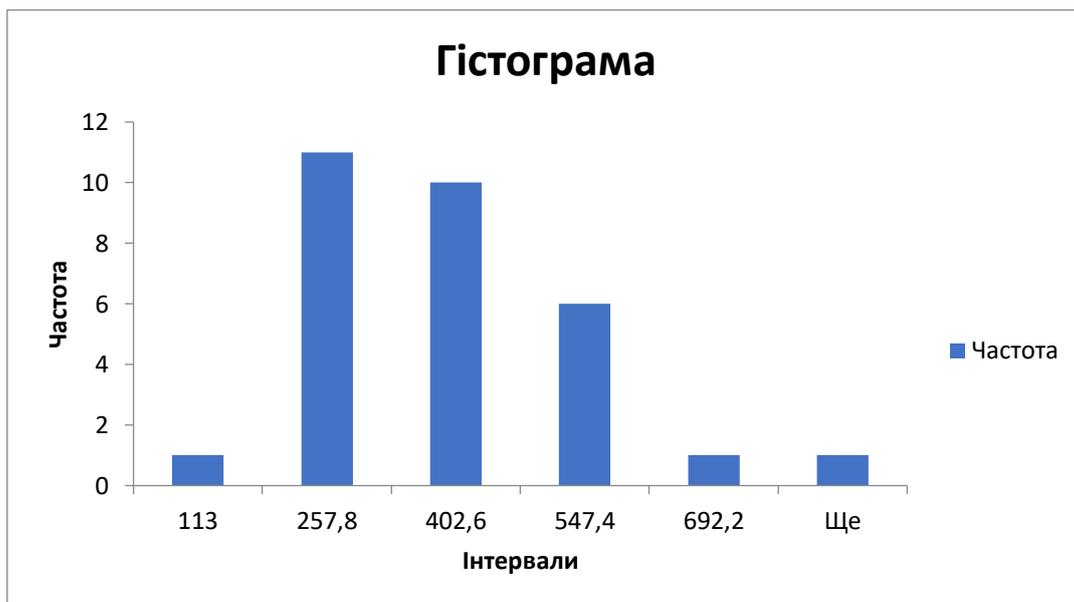
У вікні необхідно як мінімум вказати: *вхідний інтервал* і *ліву верхню комірку вихідного інтервалу*. Після натискання кнопки *ОК*:

- автоматично розраховані інтервали значень;

- підраховано кількість значень із зазначеного масиву даних, що потрапляють у кожний інтервал (побудовано таблицю частот);

Інтервали	Частота
113	1
257,8	11
402,6	10
547,4	6
692,2	1
Ще	1

- якщо поставлена галочка навпроти пункту *Виведення графіка*, то разом із таблицею частот буде виведено гістограму.



Як видно з малюнка, перший інтервал включає тільки одне мінімальне значення 113 (точніше, включені всі менші значення або рівні мінімальному). Якби в масиві було 2 або більше значення 113, то перший інтервал потрапила б відповідна кількість чисел (2 або більше).

Другий інтервал включає значення більше 113 і менше або рівні 257,8. Можна перевірити, що таких значень 11. Передостанній інтервал від 547,4 (не включаючи) до 692,2 (включаючи) містить 1

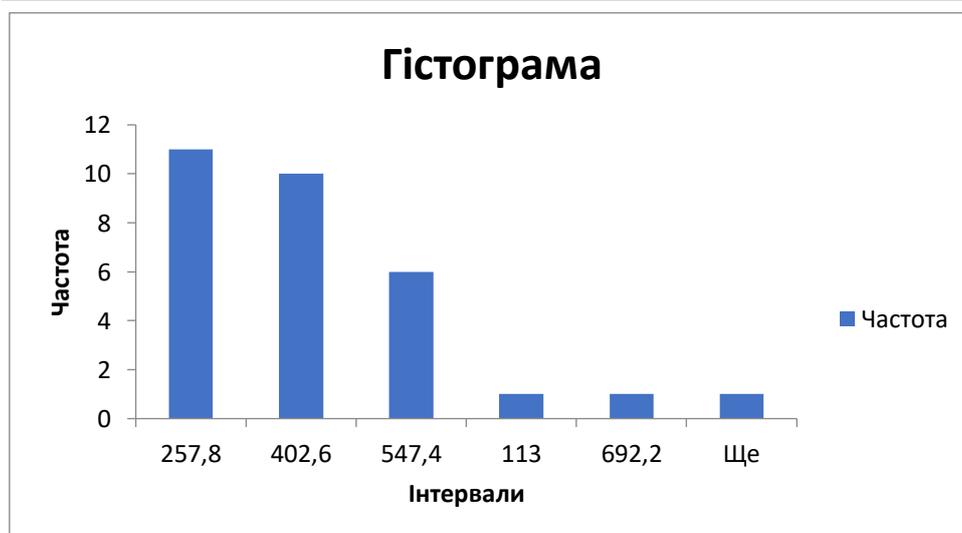
значення. Останній інтервал (з дивною назвою *ще*) містить значення більше 692,2 (не включаючи). Таких значень лише одне – максимальне значення в масиві (837).

Розміри інтервалів однакові та рівні 144,8. Це значення можна отримати так: $=(\text{МАКС}(\text{Вихідні_дані})-\text{МІН}(\text{Вихідні_дані}))/5$ де *Вихідні_дані* – іменований діапазон, що містить наші дані.

Чому 5? Справа в тому, що кількість інтервалів гістограми (кишень) залежить від кількості даних і для його визначення часто використовується формула \sqrt{n} , де n – кількість даних у вибірці. У нашому випадку $\sqrt{n} = \sqrt{30} = 5,48$ (всього 5 повноцінних кишень, тому що перший інтервал включає тільки значення рівні мінімальному).

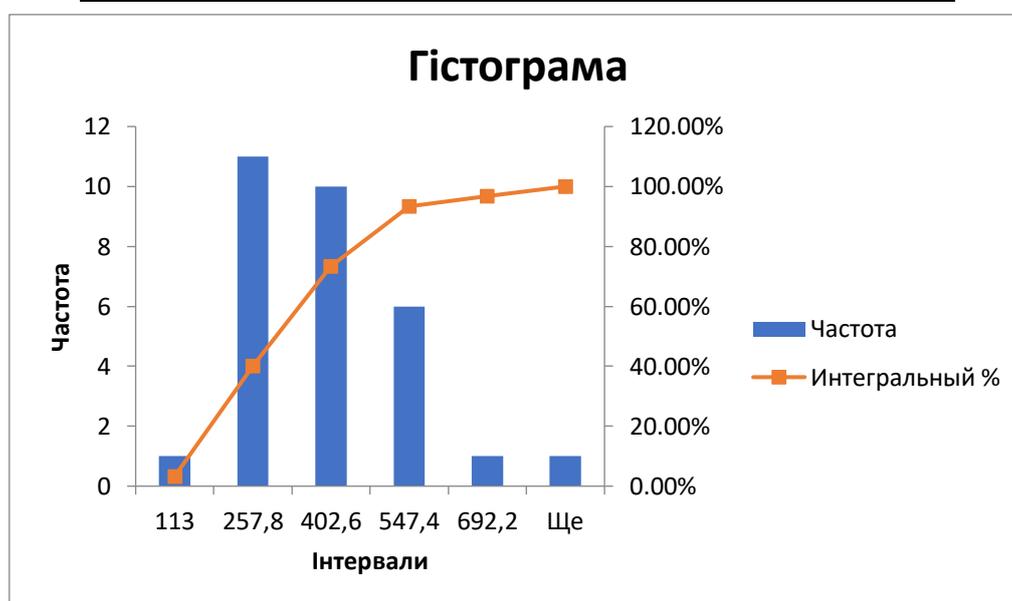
Якщо встановити галочку навпроти поля *Парето* (відсортована гістограма), то до таблиці з частотами буде додано таблицю з відсортованими за зменшенням частот.

Інтервали	Частота	Інтервали	Частота
113	1	257,8	11
257,8	11	402,6	10
402,6	10	547,4	6
547,4	6	113	1
692,2	1	692,2	1
Ще	1	Ще	1



Якщо встановити галочку навпроти поля *Інтегральний відсоток*, то до таблиці з частотами буде додано стовпець з наростаючим підсумком у % від загальної кількості значень у масиві.

<i>Інтервали</i>	<i>Частота</i>	<i>Інтегральний %</i>
113	1	3,33%
257,8	11	40,00%
402,6	10	73,33%
547,4	6	93,33%
692,2	1	96,67%
Ще	1	100,00%



Подібно до того, як описові дослідження забезпечують основу для формулювання гіпотез та планування подальших досліджень, описова статистика допомагає зрозуміти отримані дані та забезпечує базис для їхнього подальшого аналізу. Кожен історик повинен вміти інтерпретувати та обчислювати основні статистичні параметри, для того щоб розуміти взаємозв'язок даних, отриманих у випадковій вибірці, з даними генеральної сукупності, критично оцінювати дані, що публікуються, і представляти результати своїх власних досліджень належним чином.

РОЗДІЛ 4.

АНАЛІЗ ВЗАЄМОЗВ'ЯЗКІВ

4.1. Принципи функціональної залежності. Метод найменших квадратів. Кореляція. Регресія. Детермінація. Рангова кореляція

Аналіз взаємозв'язків, властивих досліджуваним процесам і явищам, є найважливішим завданням історичних досліджень. У тих випадках, коли йдеться про явища й процеси, що володіють складною структурою і різноманітним властивим їм зв'язком, такий аналіз є складним завданням. Внести сюди кількісну визначеність допомагає кореляційний зв'язок, спрямований на визначення тісноти взаємозв'язку ознак і рівня впливу різних чинників на об'єкт, що вивчається. Констатувати наявність зв'язку між ознаками дозволяють аналітичні групування, але вони не дають можливості кількісно висловити силу взаємодії однієї ознаки з іншою (парна кореляція) чи з сукупністю ознак (множинна кореляція).

Перш ніж приступати безпосередньо до кореляційного аналізу, треба перевірити правомірність його застосування – чи його результати реально відображатимуть історичну картину. Ознаки, що досліджуються методом кореляції, мають бути нормально розподілені та лінійно залежні між собою.

Ознака має властивість нормальності, якщо його значення симетрично розподіляються від «центру», яким вважається його середня арифметична величина. Найпростіше перевірити нормальність розподілу графічним методом. Графік нормально

розподіленої ознаки має дзвоноподібний вид з центром, що збігається зі значенням середнього арифметичного. В історії серед ознак, що характеризують розвиток суспільства, немає суворої нормальності розподілу. Практика використання математичних методів у суспільних науках довела доцільність відносити до нормальних розподілів ознаки з незначною мірою порушеною симетрією, з перекосами в той чи інший бік, з центром, що збігається не зі значенням середньої арифметичної, а перенесеним у максимальне значення ознаки. До нормальних можна зараховувати й графіки V-подібної форми та «перекинуті дзвони».

У статистиці розроблено багато методів аналізу розподілу значень ознак, але, по-перше, вони пов'язані зі значним обсягом обчислень; по-друге, в історичній науці поки немає необхідності отримання точних даних про особливості розподілу тих чи інших ознак. Це означає, що для перевірки правомірності використання низки математичних методів, зокрема кореляційного аналізу, в історичних роботах достатньо наближеного встановлення якості нормальності, яким виступає графік. Властивість лінійності у вивченні взаємозв'язку ознак також є необхідною попередньою умовою використання багатьох математичних методів. Лінійна залежність між двома ознаками характеризується умовою, у якій зі зміною на одиницю значень однієї ознаки змінюються у той чи інший бік значення другої. Перевірка форми залежності проводиться з допомогою графічного методу. У системі координат двох ознак точками відзначаються наявні дані. Якщо простір точок має вигляд прямої лінії, можна цю залежність характеризувати як лінійну, незалежно від напрямку точкового скупчення. Так само, як і нормальності, строгої лінійності в історії не існує. Достатньо наближеного виконання цієї властивості без залучення складніших спеціальних методик.

Отже, насамперед необхідно встановити наявність взаємозв'язків між ознаками та їхній характер. Після цього виникає питання про тісноту взаємозв'язків і ступінь впливу різних факторів (причин) на результат, що цікавить дослідника. Якщо ознаки та властивості об'єктів, що вивчаються, можуть бути

вимірянні й виражені кількісно, то аналіз взаємозв'язків може вестися на основі застосування математичних методів. Використання цих методів дозволяє перевірити гіпотезу про наявність чи відсутність взаємозв'язків між тими чи іншими ознаками, що висувуються на основі змістовного аналізу. Далі лише за допомогою математичних методів можна встановити тісноту й характер взаємозв'язків або виявити силу (ступінь) впливу різних факторів на результат.

Найбільш розробленими в математичній статистиці методами аналізу взаємозв'язків є кореляційний та регресійний аналіз. Але перш ніж переходити до їхньої характеристики, зупинимося на питанні про характер і форму тих взаємозв'язків, які притаманні об'єктивним явищам природи та суспільства.

Функціональна залежність двох кількісних ознак чи змінних полягає у тому, що кожному значенню однієї змінної завжди відповідає одне певне значення іншої змінної. Наприклад, при будівництві залізниць на кілометр колії припадає цілком певна кількість укладених рейок. Тому розглядаючи статистичні дані за такими кількісними ознаками: y – довжина прокладеної залізничної колії (в км), x – кількість витраченого на будівництві рейкового прокату (в тонах), ми матимемо справу з функціональною залежністю певного виду.

Функціональне співвідношення між ознаками зручно подати у наочній графічній формі, розглядаючи числові дані як координати точок у прямокутній системі координат. Графічним зображенням згаданої вище функціональної залежності (отриманої шляхом з'єднання безперервною лінією точок) служить пряма лінія. Така залежність називається прямою пропорційною залежністю. Її аналітичним виразом є рівняння $y = kx$, де k – коефіцієнт пропорційності. Пряма пропорційна залежність є окремим випадком лінійної залежності, яка характеризується рівнянням $y = kx + b$. Графічним зображенням лінійної залежності також є пряма лінія.

Лінійна залежність є найпростішою і у певному сенсі універсальною формою зв'язку багатьох явищ. Її універсальність у тому, що складніші залежності часто можна розглядати «у першому наближенні» як лінійні. Тут ми підходимо до з'ясування ролі функціональних залежностей в аналізі статистичних зв'язків. Безпосередньо функціональні залежності у чистому вигляді нечасто зустрічаються у суспільних явищах. Зв'язки зазвичай мають набагато складніший характер. Проте їхній вербальний опис є часто складним, та й недоцільним. Тому їх розглядають як такі, що відповідають тим чи іншим видам функціональної залежності. Найпростішою формою функціонального зв'язку є лінійна залежність, яка широко використовується у регресійному та особливо в кореляційному аналізі. Гіпотеза про лінійний зв'язок між досліджуваними ознаками набула широкого поширення в аналізі взаємозв'язків. А в тому випадку, коли результати застосування гіпотези про лінійну залежність виявляються невдалими чи є вагомі підстави заперечувати існування лінійного зв'язку, використовують складніші функціональні залежності.

Назвемо найбільш уживані форми функціональної залежності, що застосовуються у статистичному аналізі. Якщо пряма лінія не відповідає характеру даних, що використовуються, можна використовувати параболу. Аналітичний вираз її має загальний вигляд $y = ax^2 + bx + c$. Наявність у цьому рівнянні члена ax^2 є найпростішою формою обліку нелінійності.

У тому випадку, коли ми маємо справу із загасанням зростання чи падіння, зручно використовувати гіперболічні $y = \frac{1}{x}$ чи логарифмічні $y = \log_a x$ залежності. Процеси демографічного та економічного зростання описуються експоненційними залежностями $y = e^x$.

Підбір відповідної функціональної залежності на основі графічного та логічного аналізу є важливим етапом дослідження взаємозв'язків, особливо в тих випадках, коли лінійний зв'язок виявився неприйнятним.

Розглянемо особливості *статистичних (кореляційних) залежностей*. Функціональна залежність між ознаками передбачає

їхню ізольованість, вона діє, так би мовити, «за інших рівних умов». У суспільному житті такі ситуації бувають дуже рідко. Як правило, вплив однієї змінної (причини) на іншу (наслідок) відбувається не ізольовано від інших факторів, а таким чином, що на зв'язок безпосередньо чи опосередковано впливають багато інших факторів. Це вже залежність особливого вигляду. Для опису та вивчення такого роду залежностей у науці використовується поняття статистичного, чи кореляційного, зв'язку.

На відміну від функціональної залежності, коли кожному значенню однієї ознаки завжди відповідає певне значення іншого, при статистичній залежності одному й тому значенню однієї ознаки можуть відповідати різні значення іншої. Це відбувається через те, що при статистичній залежності зв'язок встановлюється між ознаками (двома, трьома тощо), які змінюються не лише через взаємодію між собою, а й під впливом безлічі різних неврахованих чинників. При цьому зв'язок між ознаками існує і проявляється в кожному конкретному випадку лише у тенденції, «у середньому». Тому тут можна встановити наявність взаємозв'язку та визначити його кількісну міру не на основі поодиноких спостережень, а лише стосовно певної сукупності об'єктів, тобто в середньому по відношенню до тих чи інших масових об'єктів чи явищ. Кількісні показники, що характеризують ці об'єкти в джерелознавстві та в статистиці, називаються масовими даними.

Вияснимо, в чому ж полягають завдання аналізу статистичних зв'язків. Аналіз статистичного, чи кореляційного, зв'язку передбачає, по-перше, виявлення форми зв'язку, а по-друге, оцінку тісноти зв'язку. Перше завдання вирішується методами регресійного аналізу, друге – методами кореляційного аналізу.

Регресійний аналіз зводиться до опису статистичного зв'язку за допомогою відповідної функціональної залежності. Метод найменших квадратів – математичний підхід для оцінки параметрів регресійних моделей на основі експериментальних даних, що містять випадкові помилки. Якщо дані відомі з деякою похибкою, замість невідомого точного значення параметра моделі використовується наближене. Тому параметри моделі мають бути

розраховані так, щоб мінімізувати різницю між експериментальними даними та теоретичними (обчисленими за допомогою запропонованої моделі). Мірою неузгодженості між фактичними значеннями та значеннями, оціненими моделлю у методі найменших квадратів, служить сума квадратів різниць між ними, тобто $\sum_{i=1}^n (y'_i - y_i)^2$, де y'_i – оцінка, отримана за допомогою моделі, y_i – фактично спостережуване значення. Очевидно, що найкращою буде та модель, яка мінімізує цю суму. Найважливішим застосуванням методу найменших квадратів у аналізі даних є лінійна регресія, де параметри регресійної моделі обчислюються таким чином, щоб сума квадратів відстаней від лінії регресії до фактичних значень даних була мінімальною.

Кореляційний аналіз дозволяє оцінювати тісноту зв'язку за допомогою спеціальних показників, причому вибір їх залежить від виду функціональної залежності, придатної для адекватного опису статистичного взаємозв'язку.

Як зазначалося, найбільш поширеною у вивченні зв'язків є гіпотеза про лінійну залежність. Відповідні їй методи кореляційного та регресійного аналізу найповніше розроблені в математичній статистиці. Перш ніж перейти до викладу цих методів, зупинимося на двох загальних питаннях, що належать до кореляційного та регресійного аналізу.

Одне з важливих питань, що виникають у вивченні зв'язків, – встановлення «напрямку» залежності. Нехай для простоти розглядається зв'язок між двома ознаками y і x . Яку з цих ознак слід вважати залежною змінною (результативною), а яку незалежною змінною (або факторною)?

Першочергове значення у вирішенні цього питання має змістовний аналіз. Припустимо, ми розглядаємо зв'язок між продуктивністю праці робітників і стажем їхньої роботи. Очевидно, результативною ознакою слід визнати продуктивність праці, а факторною – стаж робітника. Втім, не завжди «напрямок» зв'язку проявляється настільки очевидно. Тому під час вирішення питання про вибір результативної ознаки на першому плані виступає постановка змістовної проблеми, для дослідження якої використовується вивчення взаємозв'язків. Наприклад,

встановлюючи «напрямок» зв'язку між такими ознаками, як прибутковість підприємств та їхню енергоозброєність, ми повинні виходити з того, що ж ми хочемо встановити насправді: вплив впровадження нової техніки та технологій на прибутковість підприємств або потенційні можливості підприємств у оволодінні передовою технікою та технологією. У першому випадку результативною ознакою можна вважати прибутковість, у другому – енергоозброєність.

Далі кореляційний та регресійний аналіз можуть застосовуватися і дають цілком коректні результати при дотриманні певних умов. Це однорідність вихідних даних, незалежність окремих значень ознаки один від одного й нормальність розподілу ознак, що вивчаються.

Першовідкривачем кореляційного методу є французький натураліст і анатом Жорж Кюв'є (George Cuvier, 1769-1832). Закон Кореляції виведено ним засобами порівняльної анатомії. Кюв'є зрозумів, що органи одного організму відповідають один одному та його загальним умовам існування.

Отже, кореляція – показник, який відображає взаємну залежність двох чи більше величин. При цьому величини повинні вибиратися випадково, а залежність може визначатися або збігом, або відношенням причинності. Необхідно з'ясувати, чи кореляція не є хибною, тобто заснованою на збігові. Для цього запроваджується ще одна нова випадкова величина. Тільки при зміні значення однієї величини, що тягне за собою неминучу систематичну зміну значення іншої величини, кореляція вважається встановленою. Така зміна може бути виражена у вигляді коефіцієнта кореляції або кореляційного відношення. Коефіцієнт кореляції показує, наскільки тісно дві змінні пов'язані між собою.

Для візуального виявлення взаємозв'язку між кількісними змінними корисно будувати діаграми розсіювання (scatterplot). Для побудови такого графіку по горизонтальній осі (Ox) відкладається одна змінна, по вертикальній (Oy) інша. При цьому кожному

об'єкту на діаграмі відповідає точка, координати якої дорівнюють значенням пари вибраних для аналізу змінних.

Для того щоб визначити тісноту зв'язку між двома ознаками, слід вирахувати так званий парний лінійний коефіцієнт кореляції, що розраховується за формулою:

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

де x_i, y_i - значення ознак x та y для i -го об'єкта; n - число об'єктів; x, y - середні арифметичні значення ознак x та y . Лінійний коефіцієнт кореляції може приймати значення від -1 до $+1$, причому чим ближче величина коефіцієнта кореляції до граничних значень, тим тіснішим буде взаємозв'язок між ознаками. У тому випадку, якщо коефіцієнт дорівнює нулю, лінійний зв'язок між ознаками буде відсутній. Пряма функціональна залежність матиме місце, якщо коефіцієнт кореляції дорівнює $+1$ (відповідно, обернена функціональна залежність, якщо коефіцієнт рівний -1).

При аналізі історичних подій дослідники працюють переважно з якісними ознаками, різновидом яких виступають альтернативні (такі, що приймають лише два значення). Коли достатньо отримати орієнтовне уявлення про тісноту зв'язку між ознаками, можна обійтися без громіздких обчислень, звернувшись до коефіцієнта збігу символів. Метод, запропонований німецьким психіатром Г.Т. Фехнером (1801–1887 рр.), заснований на порівнянні значень ознак зі своїми середніми величинами. Якщо значення ознаки більше за його середню - воно фіксується знаком «+», якщо менше - знаком «-».

Не всяку якісну ознаку можна перетворити на альтернативну або замінити кількісною. Досягти мети виявлення та вимірювання впливу якісних ознак допомагає *коефіцієнт рангової кореляції*. Система ранжирування набула широкого поширення в історичних дослідженнях. Суть її полягає в попередній експертній оцінці варіантів якісної ознаки та присвоєння їм кількісного еквівалента, виходячи зі ступеня їхньої інтенсивності. Ранжувати ознаки, що вивчаються, обов'язково слід у тому самому порядку: або по

висхідній, або по низхідній лінії. Ранжуванню піддаються як кількісні, так і якісні ознаки. Коефіцієнт кореляції рангів може бути обчислений і для вивчення взаємозв'язку між якісним і кількісним ознаками. Ранги найчастіше позначаються порядковими числівниками – 1, 2, 3... Міру взаємозв'язку між парою ознак, кожна з яких проводить ранжування досліджуваної сукупності об'єктів, показує коефіцієнт рангової кореляції. Одну із формул коефіцієнтів кореляції рангів запропонував англійський психолог Ч. Спірмен (1863–1945 рр.).

Коефіцієнт рангової кореляції Спірмена вимірюється та інтерпретується так само, як і інші кореляційні коефіцієнти. При збігу ранжованих рядів за обома ознаками, що розглядаються, коефіцієнт прийме значення 1, що говорить про максимально тісний прямий зв'язок. Якщо об'єкти в одному ранжованому ряду прямо протилежні рангам другої ознаки, то в наявності максимально тісний зворотний зв'язок. В обох випадках обчислення коефіцієнта не є потрібним, достатньо проаналізувати взаємне розташування рангів. У історичних дослідженнях використовуються інші коефіцієнти рангової кореляції (коефіцієнт Кендалла, коефіцієнт конкордації та інші), але загальна теорія статистики рекомендує користуватися коефіцієнтом кореляції рангів Спірмена. Він менш трудомісткий та досить представницький.

Однак, часто необхідно не тільки оцінити тісноту зв'язку між ознаками, що вивчаються, але і визначити той ступінь, з яким одна ознака впливає на іншу. У цьому випадку використовується коефіцієнт детермінації, що визначає відсоткову частку змін, що відбуваються під впливом факторної ознаки, у загальній мінливості результативної ознаки:

$$D = r^2 \cdot 100\%,$$

де r – коефіцієнт кореляції.

Регресійний аналіз являє собою сукупність методів математичної статистики, які дозволяють визначити форму зв'язку

між результативною та факторною ознаками, встановленою кореляційним аналізом. Кореляційний зв'язок описується рівнянням регресії за допомогою відповідної функції. Найпростіше рівняння лінійної регресії: $y = ax + b$ де x – факторна ознака; y – результативна ознака; a і b – параметри рівняння, які можна знайти методом найменших квадратів.

Регресійний аналіз не використовується для визначення наявності зв'язку між змінними через те, що наявність такого зв'язку та є передумовою для застосування аналізу. Лінійна регресія досить добре працює у ряді простих завдань. До її переваг відноситься простота алгоритму й висока швидкість. Вона має лише один недолік – непристосованість до вирішення суто нелінійних завдань.

Кореляційний аналіз виявляє структуру взаємозв'язків ознак, що характеризують явище, яке вивчається, або процес. Але кореляційний аналіз не здатний пояснити, чим зумовлена саме така структура зв'язків. Відповісти це питання дозволяють методи факторного аналізу.

Факторний аналіз поєднує методи аналізу структури безлічі ознак, що характеризують явища, що вивчаються, і процеси, а також виявлення узагальнених факторів. В його основі лежить положення про те, що кореляційні зв'язки між великим числом показників, що спостерігаються, визначаються існуванням меншого числа гіпотетично спостережуваних показників або факторів. Пояснення безлічі вихідних ознак через невелику кількість загальних факторів здійснюється стиском інформації, що міститься у вихідних корельованих ознаках.

Основними характеристиками факторного аналізу є факторні навантаження та факторні ваги.

Факторне навантаження – це значення коефіцієнтів кореляції кожної з вихідних ознак із кожним із виявлених факторів. Чим тісніший зв'язок даної ознаки з аналізованим фактором, тим вище значення відповідного факторного навантаження. Позитивний знак факторного навантаження вказує на прямий зв'язок даної ознаки з фактором, а від'ємний знак – зворотний. Якщо значення

факторного навантаження близьке до нуля, це свідчить тому, що цей чинник мало впливає на дану ознаку. Дані про факторні навантаження дозволяють судити про вибір вихідних ознак, що відображають той чи інший фактор, і про відносну частку окремих ознак у структурі кожного фактора.

Факторні ваги – це кількісні значення виділених чинників кожного з n наявних об'єктів. У об'єктів з великими значеннями факторних ваг є великий ступінь прояву властивостей, властивих даному фактору. Фактори визначаються як стандартизовані показники із середнім арифметичним значенням 0 і середнім квадратичним відхиленням 1. Внаслідок цього додатні факторні ваги відповідають тим об'єктам, які характеризуються ступенем прояву властивостей більше середньої, а від'ємні факторні ваги відповідають тим об'єктам, у яких ступінь прояву властивостей менше середнього. Об'єкти ранжуються в межах кожного фактора, виходячи з даних про факторні ваги. А їх, у свою чергу, можна розглядати як значення індексу, що характеризує рівень розвитку об'єктів у аспекті, що розглядається.

Відповідно, факторні ваги можуть стати основою класифікації об'єктів, що досліджуються. Створення багатовимірної типології з урахуванням факторного аналізу є особливо ефективним у випадку, якщо є велика кількість ознак, що характеризують сукупність об'єктів, та їхній змістовий відбір становить значні труднощі. У такій ситуації необхідне «стиснення» інформації, після чого проводиться класифікація за будь-яким із виділених факторів.

4.2. Застосування інструментів EXCEL для аналізу взаємозв'язків

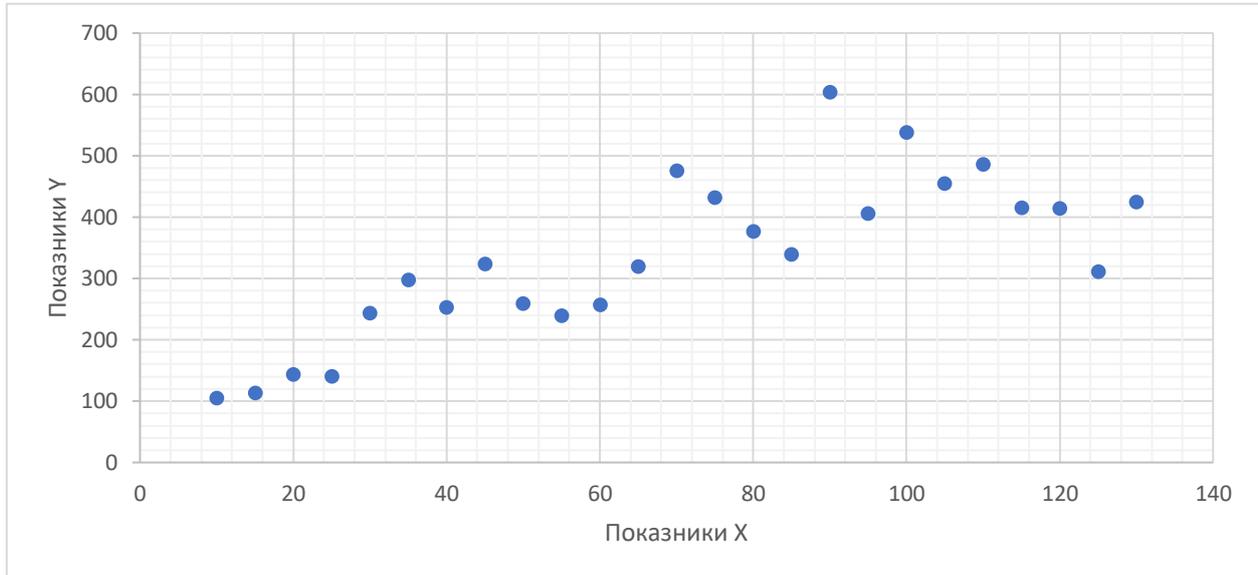
Регресія дозволяє прогнозувати залежну змінну виходячи зі значень фактора. У MS EXCEL є безліч функцій, які повертають не лише нахил і зсув лінії регресії (нагадаємо, що формула лінії

регресії має вигляд $y = kx + b$ де k - нахил, b - зсув), що характеризують лінійний взаємозв'язок між факторами, а й регресійну статистику. Тут розглянемо просту лінійну регресію, тобто прогнозування з урахуванням одного чинника.

Нехай у нас є масив даних, що представляє собою значення двох змінних X та Y . Причому значення змінної X ми можемо довільно задавати (контролювати) і використовувати цю змінну для передбачення значень залежної змінної Y . Таким чином, випадковою величиною є лише змінна Y .

№ вимірювання	показник X	показник Y
1	10	104,3807
2	15	112,9988
3	20	143,2934
4	25	140,4546
5	30	243,4168
6	35	296,8972
7	40	253,1406
8	45	323,2403
9	50	258,5827
10	55	239,0081
11	60	256,7887
12	65	319,4259
13	70	475,6097
14	75	431,7905
15	80	376,8099
16	85	338,701
17	90	603,1855
18	95	406,0964
19	100	538,1194
20	105	454,9282
21	110	486,159
22	115	415,0824
23	120	413,8162
24	125	311,0506
25	130	424,7028

Побудуємо діаграму розсіювання.



Наведена вище діаграма розсіювання свідчить про можливий лінійний взаємозв'язок між Y від X : очевидно, що точки даних в основному розташовуються вздовж певної прямої лінії.

Втім, наявність навіть такого очевидного лінійного взаємозв'язку не може бути доказом про наявність причинного взаємозв'язку змінних. Наявність причинного взаємозв'язку може бути доведено виходячи з лише аналізу наявних вимірів.

Розглянемо інструмент *Регресія*, що входить до надбудови *Пакет Аналізу*. У діалоговому вікні *Аналіз даних* виберіть інструмент *Регресія*. Після натискання кнопки *ОК* буде виведено інше діалогове вікно.

У цьому вікні потрібно вказати:

- *Вхідний інтервал Y*: посилання на масив значень змінної Y. Посилання можна вказати із заголовком. У цьому випадку, при виведенні результатів, надбудова використовує Ваш заголовок (для цього у вікні потрібно встановити галочку *Мітки*);
- *Вхідний інтервал X*: посилання на значення змінної X. У разі множинної регресії (кілька змінних X) потрібно вказати всі стовпці зі значеннями X. У разі множинної регресії посилання рекомендується робити на діапазон із заголовками (у вікні потрібно встановити галочку *Мітки*);
- *Константа-нуль*: якщо галочка встановлена, то надбудова підбирає лінію регресії, що проходить через точку $Y = 0$ (зсув дорівнює 0);
- *Рівень надійності*: Це значення використовується для побудови довірчих інтервалів для нахилу та зсуву. Рівень надійності = $1 - \alpha$. Якщо галочка не встановлена чи встановлена, але рівень значущості = 95%, то надбудова все одно розраховує межі довірчих інтервалів, причому дублює їх. Якщо галочка встановлена, а рівень надійності відмінний від 95%, то

розраховуються 2 довірчі інтервали: один для 95%, інший для введеного значення;

- *Вихідний інтервал*: діапазон комірок, куди будуть розміщені результати обчислень. Достатньо вказати ліву верхню комірку цього діапазону;
- *Залишки*: буде обчислено залишки моделі, тобто різниця між спостереженими та передбаченими значеннями Y_i для всіх спостережень n ;
- *Стандартизовані залишки*: Вищезазначені значення залишків будуть поділені на значення їхнього стандартного відхилення;
- *Графік залишків*: Буде побудовано точкову діаграму: значення залишків для всіх значень X_i ;
- *Графік підбору*: Буде побудована точкова діаграма: точки даних $(X; Y)$ та лінія регресії;
- *Графік нормальної ймовірності*: Буде побудована точкова діаграма під назвою *Графік нормального розподілу*. По суті – це графік значень змінної Y , відсортованих за зростанням.

Звіт, сформований надбудовою, складається з наступних розділів:

Розділ «Регресійна статистика»:

- *Множинний R*. У разі простої лінійної регресії – це Коефіцієнт кореляції, функція КОРРЕЛ()
- *R-квадрат*. У разі простої лінійної регресії – це коефіцієнт детермінації, функція КВПРСОН()
- *Нормований R-квадрат*. Коефіцієнт детермінації;
- *Стандартна помилка*. Стандартна помилка регресії;
- *Спостереження*. Кількість значень Y .

Розділ «Дисперсійний аналіз»:

- *df* – ступені свободи (Degrees of Freedom).
- *SS* – сума квадратів (Sum of Squares)
- *MS* – SS/df (MSR та MSE)
- *F* – значення статистики F_0 (MSR/MSE)
- *Значення F* – *p*-значення, функція F.РАСП.ПХ()

Інші результати:

- *Коефіцієнти*: оцінка параметрів моделі k та b .
- *Стандартна помилка*: Стандартні помилки вищевказаних статистик
- *t-статистика*: значення тестової статистики t_0 .
- *P-Значення*: перевірка значимості взаємозв'язку змінних
- *Нижні 95% та Верхні 95%*: межі довірчих інтервалів для оцінок невідомих параметрів моделі k та b .

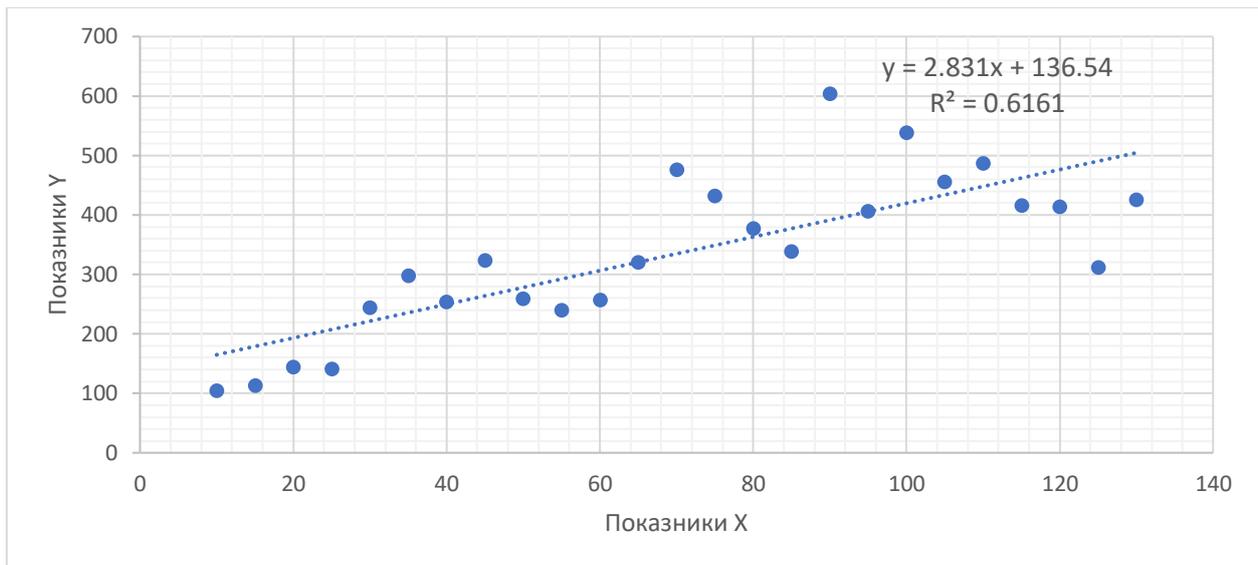
В результаті застосування інструменту *Регресія* будуть виведені такі статистичні показники:

<i>Регресійна статистика</i>	
Множинний R	0,78492512
R-квадрат	0,616107444
Нормований R-квадрат	0,599416464
Стандартна похибка	84,002975
Спостереження	25

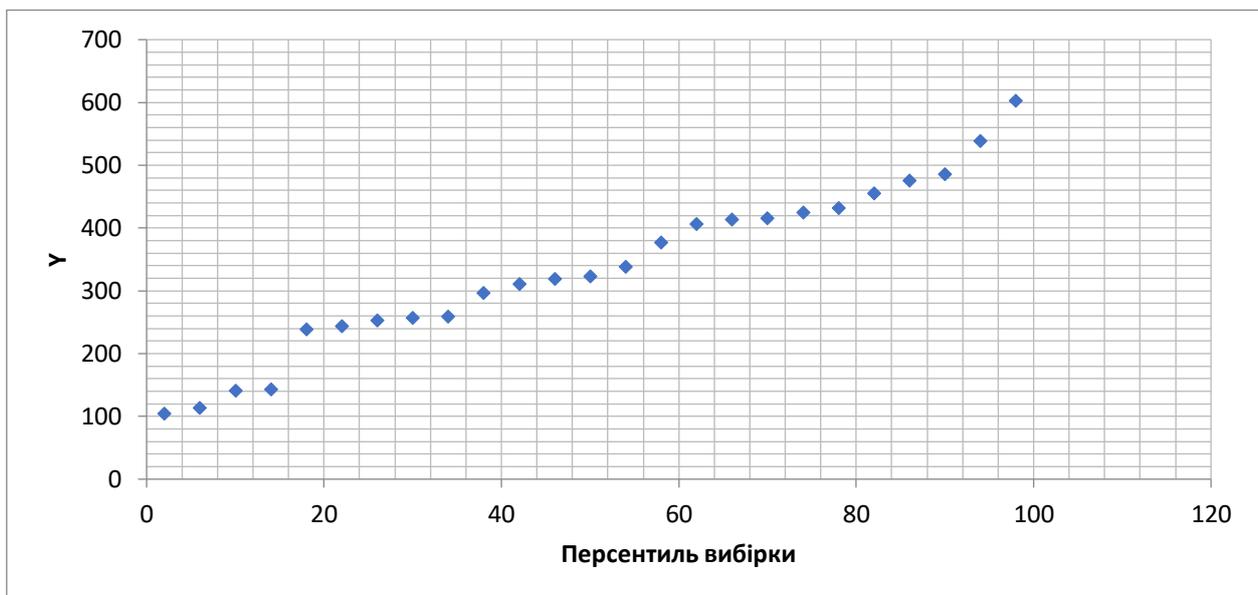
<i>Дисперсійний аналіз</i>					
	<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Значимість F</i>
Регресія	1	260473,7341	260473,7341	36,91259706	3,3859E-06
Залишок	23	162299,4956	7056,499809		
Підсумок	24	422773,2297			

	<i>Коефіцієнти</i>	<i>Стандартна похибка</i>	<i>t-статистика</i>	<i>P-Значення</i>	<i>Нижні 95%</i>	<i>Верхні 95%</i>
<i>Зсув</i>	136,5369853	36,69009458	3,721358225	0,001120796	60,63774188	212,4362287
<i>Нахил</i>	2,831002726	0,465964667	6,075573805	3,3859E-06	1,867081371	3,794924081

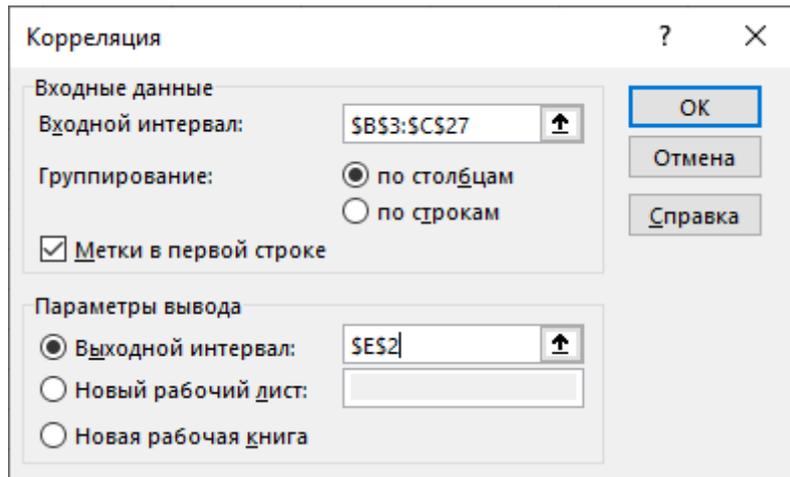
Графік підбору (точкова діаграма та лінія регресії) буде виглядати так:



Графік нормального розподілу буде таким:



Для обчислення величини парного кореляційного зв'язку можемо також використати інструмент *Кореляція*, що входить до надбудови *Пакет Аналізу*. У діалоговому вікні *Аналіз даних* виберіть інструмент *Кореляція*. Після натискання кнопки *ОК* буде виведено діалогове вікно.



Діалогове вікно містить такі поля:

- *Вхідний інтервал*: потрібно ввести посилання на діапазон з вихідними даними для 2-х змінних;
- *Групування*: як правило, вихідні дані вводяться в 2 стовпці;
- *Мітки в першому рядку*: якщо встановлена галочка, то *Вхідний інтервал* повинен містити заголовки стовпців. Рекомендується встановлювати галочку, щоб результат роботи *Надбудови* містив інформативні стовпці
- *Вихідний інтервал*: діапазон комірок, куди будуть розміщені результати обчислень. Достатньо вказати ліву верхню комірку цього діапазону.

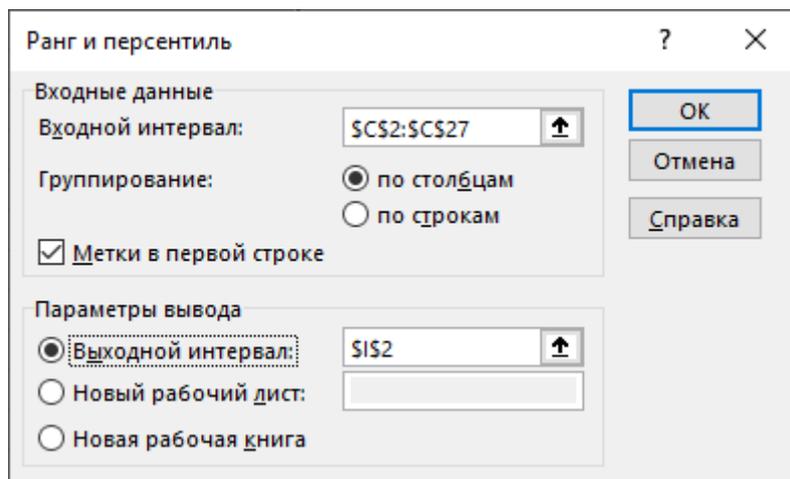
В результаті роботи інструменту отримаємо таку таблицю:

	показн X	показн Y
показн X	1	
показн Y	0,784925	1

Для визначення кореляційного зв'язку між масивами кількісних та якісних даних останнім присвоюються ранги. Більше того, так поступають і в тих випадках, коли при визначенні кореляційного зв'язку між масивами з кількісними даними різниця між значеннями даних одного або обох масивів є несуттєвою для отримання результатів дослідження. Ранжування даних – це впорядкований запис показників у порядку зростання або спадання ознаки, що вивчається. Ранжування зазвичай передують групуванню, яке є об'єднанням одиниць, що вивчаються в сукупності якісно однорідної групи за якою-небудь ознакою

(причому ознака не повинна бути випадковою). Групування застосовується в аналізі для розкриття змісту середніх підсумкових показників та впливу окремих одиниць на їхні середні показники.

Для ранжування кількісних даних використовується інструмент *Ранг та Персентиль*, що входить до надбудови *Пакет Аналізу*. У діалоговому вікні *Аналіз даних* виберіть інструмент *Ранг та Персентиль*. Після натискання кнопки *ОК* буде виведено діалогове вікно.



Якщо посилання на масив вказати разом із заголовком, потрібно встановити галочку в полі *Мітки* в першому рядку. Вказувати заголовки зручно, коли аналізується одразу кілька масивів даних. Втім, можна вказати посилання на діапазон лише з числами та не ставити галочку.

Посилання на вихідний інтервал вкажемо на поточному аркуші, наприклад, у комірці I2 (це буде верхній лівий кут таблиці результатів). Після натискання кнопки *ОК* буде виведено таблицю результатів, що містить такі стовпці:

<i>Точка</i>	<i>показн Y</i>	<i>Ранг</i>	<i>Процент</i>
17	603,1855	1	100,00%
19	538,1194	2	95,80%
21	486,159	3	91,60%
13	475,6097	4	87,50%
20	454,9282	5	83,30%
14	431,7905	6	79,10%

25	424,7028	7	75,00%
22	415,0824	8	70,80%
23	413,8162	9	66,60%
18	406,0964	10	62,50%
15	376,8099	11	58,30%
16	338,701	12	54,10%
8	323,2403	13	50,00%
12	319,4259	14	45,80%
24	311,0506	15	41,60%
6	296,8972	16	37,50%
9	258,5827	17	33,30%
11	256,7887	18	29,10%
7	253,1406	19	25,00%
5	243,4168	20	20,80%
10	239,0081	21	16,60%
3	143,2934	22	12,50%
4	140,4546	23	8,30%
2	112,9988	24	4,10%
1	104,3807	25	0,00%

Стовпці мають такий зміст.

- *Стовпець із заголовком Точка.* У ньому виведені початкові позиції значень зі стовпця «показник Y». Якщо у вхідному масиві є повтори, то кожному повтору відповідатиме своя позиція. Цей стовпець можна сформувавши за допомогою масивної формули;
- *Стовпець «показник Y».* Це просто наш вхідний масив сортований за спаданням. Його можна сформувавши формулою =НАЙБІЛЬШИЙ();
- *Стовпець Ранг.* Це стовпець рангів значень. Максимальному значенню буде надано ранг=1. Цей стовпець можна сформувавши за допомогою формули =РАНГ.РВ ();
- *Стовпець Процент.* Стовпець містить процентиль-ранг (percentile rank), який є оцінкою відносного положення значення в масиві. Для заданого значення обчислюється

скільки значень у масиві менше або дорівнює йому. Точніше – який відсоток значень масиву менший чи дорівнює йому. Зрозуміло, що з максимального значення процентиль-ранг дорівнює 0,00%, а мінімального – 100% (всі значення масиву менше або дорівнюють йому). Повтори в масиві мають той самий ранг і процентиль-ранг. Цей стовпець можна сформулювати за допомогою формули =ПРОЦЕНТРАНГ.ВКЛ ().

Незважаючи на свою назву *Ранг та Персентиль*, надбудова *Пакет аналізу* не повертає *Персентилі*. Принаймні у тому розумінні, яке закладено у функції ПРОЦЕНТИЛЬ.ВКЛ().

РОЗДІЛ 5.

АНАЛІЗ ДИНАМІЧНИХ РЯДІВ

5.1. Основні характеристики динамічного ряду. Склад динамічного ряду: тренд, сезонна та стохастична компоненти. Вирівнювання. Стандартизація та нормалізація рядів. Кореляція. Автокореляція

При вивченні динаміки в часі будь-якого явища вдаються до побудови динамічного ряду.

Динамічний ряд – це ряд однорідних статистичних величин, що показують зміну будь-якого явища в часі та розташованих у хронологічному порядку через певні проміжки часу. Числа, що становлять динамічний ряд, називаються рівнями.

Рівень ряду – розмір (величина) того чи іншого явища, досягнутий у певний період чи певний момент часу. Рівні ряду можуть бути представлені абсолютними, відносними чи середніми величинами.

Динамічні ряди поділяються на

а) *прості* (що складаються з абсолютних величин) – можуть бути:

1) *моментними* – складаються з величин, що характеризують явище на якийсь певний момент (статистичні відомості, які

зазвичай реєструються на початок або кінець місяця, кварталу, року)

2) *інтервальними* – складаються з чисел, що характеризують явище за певний проміжок часу (інтервал) – за тиждень, місяць, квартал, рік (дані про кількість народжених, померлих за рік, кількість інфекційних захворювань за місяць). Особливістю інтервального ряду є те, що його члени можна підсумовувати (при цьому укрупнюється інтервал), або подрібнювати.

б) *складні* (що складаються з відносних чи середніх величин).

Динамічні ряди можуть зазнавати перетворень, метою яких є виявлення особливостей зміни досліджуваного процесу, а також досягнення наочності.

Показники динамічного ряду є такі:

а) *рівні ряду* – величини членів ряду. Величина першого члена ряду називається початковим (вихідним) рівнем, величина останнього члена ряду – кінцевим рівнем, середня величина з усіх членів ряду називається середнім рівнем.

б) *абсолютний приріст (спад)* – величина різниці між наступним та попереднім рівнями; приріст виражається числами з додатним знаком, спад – з від'ємним знаком. Значення приросту чи спаду відображають зміни рівнів динамічного ряду за певний проміжок часу.

в) *темп зростання (зниження)* – показує відношення кожного наступного рівня до попереднього рівня і зазвичай визначається у відсотках.

г) *темп приросту (спаду)* – відношення абсолютного приросту або втрати кожного наступного члена ряду до рівня попереднього, виражене у відсотках. Темп приросту може бути обчислений також за формулою: Темп зростання – 100%

Абсолютне значення одного відсотка приросту (спаду) знаходиться з відношення абсолютної величини приросту (або спаду) до показника темпу приросту (або спаду) за той же період.

Для більш наочного вираження зростання чи спадання ряду можна перетворити його шляхом обчислення показників

наочності, що показують відношення кожного члена ряду до одного з них, прийнятого за сто відсотків.

До основних причин, які ускладнюють або унеможливають зіставлення рівнів динамічного ряду можна віднести:

- зміни одиниці виміру або підрахунків (оцінка економічної ефективності роботи установ у різних грошових еквівалентах на дані періоди – долари, купони, гривні);
- нерівномірна періодизація динаміки (кількісна – за роками, якісна – за соціально-економічними періодами, зміни пріоритетності різних типів установ);
- зміни переліку об'єктів аналізу (перехід низки установ з одного підпорядкування до іншого);
- зміни територіальних кордонів областей, районів та ін.

За наявності вищевказаних умов проблему, як правило, вирішують у процесі збирання та обробки даних або шляхом їхнього перерахунку.

Іноді динаміка досліджуваного явища представлена не у вигляді безперервно змінюваного рівня, а окремими стрибкоподібними змінами. Тоді для виявлення основної тенденції у розвитку досліджуваного явища вдаються до *вирівнювання динамічного ряду*. При цьому можуть бути використані такі прийоми:

а) *укрупнення інтервалу* – підсумовування даних за низку суміжних періодів. В результаті виходять підсумки за більш тривалі часові відтинки. Цим згладжуються випадкові коливання і чіткіше визначається характер динаміки явища;

Місяць	1	2	3	4	5	6	7	8	9	10	11	12	всього
Кількість подій	130	210	122	166	227	160	184	240	156	233	240	199	2267
Укрупнення рівнів	462		553			580			672				

б) *обчислення групового середнього* – визначення середньої величини кожного укрупненого періоду. Для цього необхідно

сумувати суміжні рівні сусідніх періодів, а потім суму поділити на кількість доданків. Цим досягається більша ясність змін у часі;

Місяць	1	2	3	4	5	6	7	8	9	10	11	12	всього
Кількість подій	130	210	122	166	227	160	184	240	156	233	240	199	2267
Групове середнє	462/3 = 154			553/3 = 184			580/3 = 193			672/3 = 224			

в) обчислення ковзного середнього – певною мірою усуває вплив випадкових коливань на рівні динамічного ряду й більш помітно відбиває тенденцію явища. При його обчисленні кожен рівень ряду замінюється на середню величину з цього рівня та двох сусідніх із ним. Найчастіше підсумовуються послідовно три члени ряду, але можна брати й більшу кількість;

Місяць	1	2	3	4	5	6	7	8	9	10	11	12	всього
Кількість подій	130	210	122	166	227	160	184	240	156	233	240	199	2267
Ковзне середнє	154												
Ковзне середнє		166											
Ковзне середнє			172										
Ковзне середнє				184									
Ковзне середнє					190								
Ковзне середнє						194							

Ковзне середнє								193				
Ковзне середнє									209			
Ковзне середнє										209		
Ковзне середнє											224	

г) *графічний метод* – вирівнювання від руки або за допомогою лінійки, циркуля графічного зображення динаміки явища, що вивчається;

д) *вирівнювання методом найменших квадратів* – один із найбільш точних способів вирівнювання динамічного ряду (дивіться попередній розділ посібника). Метод має на меті усунути вплив тимчасово діючих причин, випадкових факторів і виявити основну тенденцію в динаміці явища, спричинену впливом лише тривалих факторів. Вирівнювання проводиться по лінії, що найбільш відповідає характеру динаміки явища, що вивчається, за наявності основної тенденції до зростання або зниження частоти явища. Такою лінією є зазвичай пряма, яка найбільш точно характеризує основний напрямок змін, проте існують інші залежності (квадратична, кубічна і т.д.). Цей метод дозволяє дати кількісну оцінку виявленої тенденції, оцінити середні темпи її розвитку та розрахувати прогнозовані рівні наступного року;

е) *експоненційне згладжування (Exponential Smoothing)* – цей метод надає більшого значення останнім спостереженням, зменшуючи вагу попередніх значень експоненційно. Наприклад, коли згладжується часовий ряд, найсвіжіші дані більше впливають на прогноз, ніж старіші. Просте експоненційне згладжування (SES) добре працює для рядів без трендів або сезонних компонент, а

розширені версії, як-от Holt's Linear Trend і Holt-Winters, враховують тренд і сезонність.

є) *лінійна фільтрація* – метод полягає в застосуванні фільтрів для згладжування даних, відфільтровуючи високочастотні випадкові коливання. Один із класичних прикладів – фільтр Ходріка-Прескотта (Hodrick-Prescott Filter), який розділяє часовий ряд на трендову та циклічну компоненти.

Отже, вирівнювання динамічних рядів – це важливий інструмент для виявлення довгострокових тенденцій і полегшення аналізу часових рядів.

Сезонна компонента динамічного ряду – це частина ряду, яка відображає регулярні коливання даних, що повторюються через рівні інтервали часу (сезони), зазвичай у межах одного року. Ці коливання пов'язані з природними або соціальними факторами, які повторюються щороку в певні періоди.

Прикладом сезонної компоненти можуть бути зростання попиту на електроенергію взимку через опалення, підвищення продажів подарунків напередодні свят, коливання урожайності через сезонні цикли аграрної діяльності.

Сезонні коливання можна виявити, аналізуючи динамічний ряд і визначаючи періодичність таких змін. Важливо зазначити, що сезонна компонента є регулярною і передбачуваною на основі історичних даних. В аналізі часових рядів її часто виділяють окремо для більш точного прогнозування або аналізу трендів.

Моделі, які можуть враховувати сезонну компоненту, включають ARIMA з сезонним компонентом (SARIMA) або декомпозицію часових рядів на тренд, сезонність та залишки.

Стохастична компонента динамічного ряду – це частина ряду, яка відповідає за випадкові, непередбачувані коливання даних. Вона відображає вплив факторів, що не мають чіткої закономірності або періодичності та є результатом дії випадкових процесів. Стохастичну компоненту часто називають шумом, залишком або випадковими коливаннями. Вона може включати вплив таких факторів, як раптові зміни в економічних умовах, непередбачувані

події, як-от природні катастрофи або політичні кризи, помилки вимірювання або інші випадкові відхилення.

На відміну від сезонної та трендової компонент, стохастичні зміни не мають регулярного характеру, і їх складно або неможливо передбачити на основі минулих даних. В аналізі часових рядів стохастичну компоненту намагаються відокремити від інших компонент (сезонної та трендової), щоб зрозуміти більш регулярні тенденції або цикли в даних.

Залишкова стохастична компонента часто є предметом аналізу в таких моделях, як ARIMA, де вона моделюється як білий шум або авторегресійні випадкові процеси.

Зазначимо, що ARIMA (*AutoRegressive Integrated Moving Average*) називають статистичну модель для аналізу та прогнозування часових рядів. Вона поєднує три основні елементи:

AR (авторегресійний компонент): модель, яка припускає, що поточне значення ряду залежить від попередніх значень. Наприклад, якщо у вас є часовий ряд даних, то його поточне значення може бути виражене як лінійна комбінація попередніх спостережень.

I (інтеграція): процедура, що використовується для перетворення нестабільного (нерухомого) часового ряду в стаціонарний. Інтеграція полягає в обчисленні різниць між послідовними значеннями ряду, що дозволяє усунути тренди або інші нестабільні компоненти.

MA (ковзне середнє): модель, яка враховує вплив попередніх випадкових похибок (шуму) на поточні значення ряду. Вона моделює залежність поточного значення від попередніх залишкових значень (випадкових помилок).

Загальний вигляд моделі ARIMA(p, d, q):

p – порядок авторегресійного компонента (AR), тобто кількість попередніх значень, які враховуються.

d – ступінь інтеграції, тобто кількість різниць, необхідних для стаціонаризації ряду.

q – порядок ковзного середнього компонента (МА), що вказує, скільки попередніх помилок враховується в моделі.

Приклад застосування ARIMA є таким. Уявімо часовий ряд подій, який має сезонні та трендові коливання. Якщо цей ряд показує тенденцію до зростання або спадання, його необхідно стаціонаризувати (декомпозувати на стаціонарну серію через інтеграцію). Далі застосовуються авторегресійні та ковзні середні для моделювання і прогнозування майбутніх значень ряду на основі його історичних даних.

Важливі особливості ARIMA полягають в тому, що: а) ARIMA використовується для прогнозування нестабільних (нестационарних) рядів, які можуть мати тренди або сезонність; б) модель може бути розширена до SARIMA (Seasonal ARIMA) для врахування сезонних ефектів.

Для порівняння динамічних рядів часто використовуються процедури стандартизації та нормалізації.

Стандартизація динамічних рядів – це метод перетворення даних часових рядів таким чином, щоб вони мали спільну шкалу без зміни їхніх характеристик і структури. Зазвичай стандартизація передбачає приведення ряду до середнього значення 0 і дисперсії 1. Цей процес допомагає зробити дані порівняльними, особливо коли йдеться про ряди з різними одиницями виміру або масштабами.

Стандартизація проводиться з метою а) узгодження різних масштабів: коли порівнюються різні часові ряди або аналізується ряд з великими коливаннями, стандартизація дозволяє зробити їх взаємозамінними і зіставними; б) підвищення стабільності та точності моделей: деякі алгоритми машинного навчання, наприклад, методи регресії, кластеризації або нейронні мережі, працюють ефективніше на стандартизованих даних; в) усунення впливу різних одиниць виміру: стандартизація важлива, коли ряд містить значення в різних одиницях виміру, як-от долари, відсотки чи інші економічні показники.

Для перетворення кожного значення x_t часового ряду з метою його стандартизації використовують таку формулу:

$$z_t = \frac{x_t - \mu}{\sigma},$$

де x_t – значення часового ряду на момент часу t , μ – середнє значення ряду, σ – стандартне відхилення ряду, z_t – стандартизоване значення.

Стандартизація динамічних рядів покращує стабільність – допомагає уникнути проблем, пов'язаних із великими або дуже малими значеннями в ряді, що може заважати правильній роботі моделей. Крім того стандартизація спрощує порівняння – стандартизація дозволяє порівнювати різні часові ряди, навіть якщо вони мають різну середню величину або дисперсію.

Стандартизація застосовується в таких випадках: а) коли часові ряди мають різні одиниці виміру або різні масштаби; б) перед застосуванням методів машинного навчання, які чутливі до масштабу даних (лінійна регресія, кластеризація методом k -середніх тощо); в) коли потрібно усунути вплив дисперсії і середніх значень для виявлення прихованих закономірностей.

Нормалізація динамічних рядів – це метод перетворення даних часових рядів, метою якого є приведення значень до певного інтервалу, зазвичай $[0, 1]$. Це дозволяє масштабувати дані, зберігаючи при цьому їхню пропорційність, що важливо для порівняння або подальшого аналізу, особливо в тих випадках, коли дані мають різні одиниці виміру або великі амплітуди коливань.

Тому мета нормалізації полягає у: а) приведенні до спільного масштабу: це корисно, коли ряди мають різні величини або одиниці виміру, що робить їх складними для порівняння або обробки; б) підвищенні точності моделей: багато алгоритмів машинного навчання, наприклад, методи нейронних мереж або методи на основі відстаней (k -середніх, найближчий сусід), вимагають нормалізованих даних для коректної роботи; в) усунення впливу екстремальних значень: нормалізація допомагає зменшити вплив аномально великих або малих значень у часових рядах.

Найбільш поширений спосіб нормалізації – це мінімаксне масштабування, яке перетворює значення ряду на інтервал $[0, 1]$ за такою формулою:

$$z_t = \frac{x_t - x_{min}}{x_{max} - x_{min}},$$

де x_t – значення часового ряду в момент часу t , x_{min} – мінімальне значення в ряді, x_{max} – максимальне значення в ряді, z_t – нормалізоване значення на інтервалі $[0, 1]$.

Наприклад, якщо значення часового ряду коливаються від 100 до 5000, нормалізація призведе до того, що всі значення будуть приведені до інтервалу $[0, 1]$, де 100 стане 0, а 5000 стане 1. Це дозволяє порівнювати ряди з різними шкалами.

Нормалізація динамічних рядів має такі переваги: а) узгодження різних одиниць виміру: коли ряди мають різні одиниці (наприклад, відстань в кілометрах та час у годинах), нормалізація дає можливість порівнювати їх; б) зменшення впливу екстремальних значень: екстремальні значення більше не впливають на інтерпретацію загального ряду; в) покращення ефективності алгоритмів: алгоритми кластеризації, регресії та нейронні мережі часто вимагають, щоб дані знаходилися в однакових діапазонах.

Нормалізацію динамічних рядів доцільно застосовувати у таких випадках: а) коли необхідно порівнювати часові ряди, що мають різні масштаби; б) перед застосуванням алгоритмів, чутливих до масштабу даних (нейронні мережі, метод k -середніх, аналіз на основі відстаней тощо); в) для підготовки даних перед візуалізацією, щоб зробити діаграми зручнішими для інтерпретації.

Отже, нормалізація динамічних рядів допомагає усунути різницю в масштабах і підготувати дані до подальшого аналізу, зберігаючи при цьому їхню пропорційну структуру.

За наявності у часовому ряді тенденції та циклічних коливань значення кожного наступного рівня ряду залежать від попередніх.

Кореляційну залежність між послідовними рівнями часового ряду називають *автокореляцією* рівнів ряду.

Кількісно її можна виміряти за допомогою лінійного коефіцієнта кореляції між рівнями вихідного часового ряду та рівнями цього ряду, зрушеними на кілька кроків у часі. Цю величину називають коефіцієнтом автокореляції рівнів ряду першого порядку, оскільки він вимірює залежність між сусідніми рівнями ряду. Аналогічно можна визначити коефіцієнти автокореляції другого та вищих порядків.

Якщо найвищим виявився коефіцієнт автокореляції першого порядку, досліджуваний ряд містить лише тенденцію. Якщо найвищим виявився коефіцієнт автокореляції порядку t , ряд містить циклічні коливання з періодичністю в t моментів часу. Якщо жоден з коефіцієнтів автокореляції не є значущим, можна зробити одне з двох припущень щодо структури цього ряду: або ряд не містить тенденції і циклічних коливань, або містить сильну нелінійну тенденцію, для виявлення якої потрібно провести додатковий аналіз.

Коефіцієнт автокореляції будується за аналогією з лінійним коефіцієнтом кореляції і таким чином характеризує тісноту лише лінійного зв'язку поточного та попереднього рівнів ряду. Тому за коефіцієнтом автокореляції можна будувати висновки про наявності лінійної (чи близької до лінійної) тенденції.

Разом з тим за знаком коефіцієнта автокореляції не можна робити висновок про зростаючу або спадну тенденцію в рівнях ряду. Більшість часових рядів історичних та економічних даних містять додатну автокореляцію рівнів, проте при цьому можуть мати спадну тенденцію.

Число періодів, якими розраховується коефіцієнт автокореляції, називають лагом.

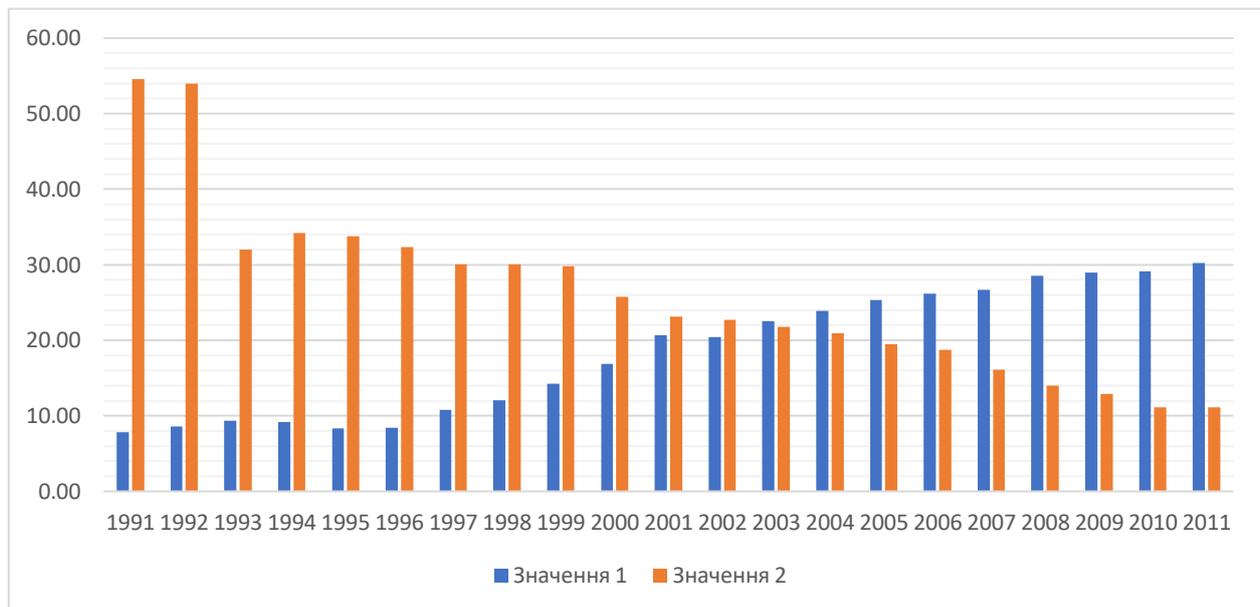
Послідовність коефіцієнтів автокореляції рівнів першого, другого й інших порядків називають *автокореляційною функцією часового ряду*. Графік залежності її значень від величини лага (порядку коефіцієнта автокореляції) називається *корелограмою*.

5.2. Застосування інструментів EXCEL для аналізу динамічних рядів

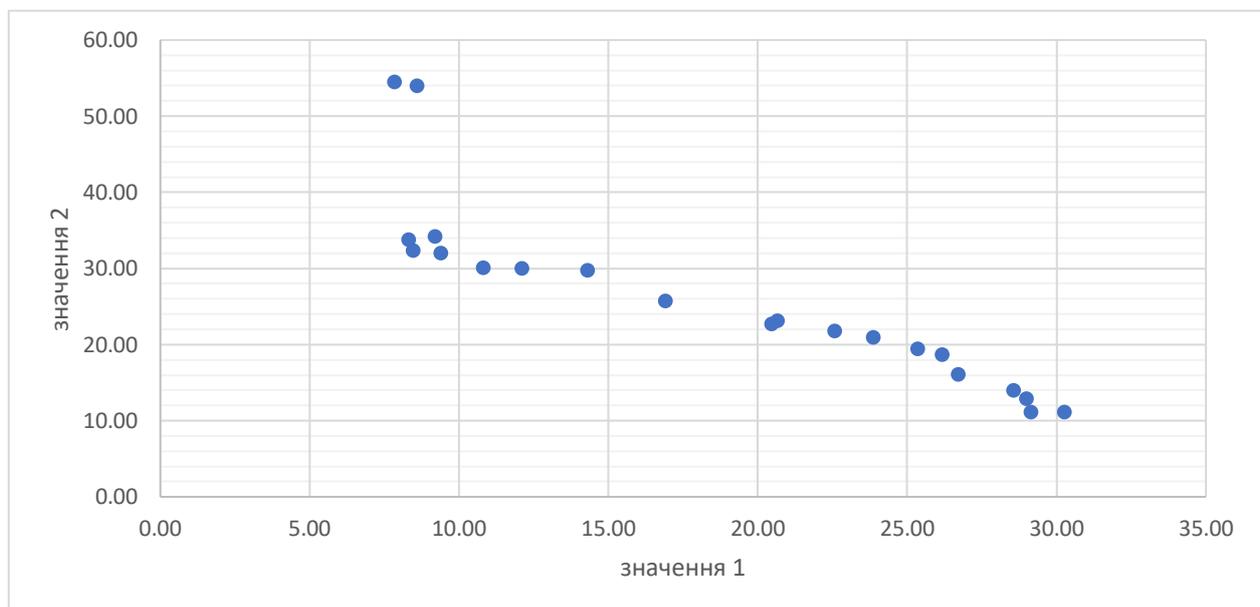
Нехай у нас є масив динамічних даних, що представляє собою кількісні значення двох явищ: середнє значення забезпеченості закладів вищої освіти України персональними комп'ютерами навчального призначення у розрахунку на 100 студентів денної форми навчання (Значення 1) та коефіцієнти варіації забезпеченості закладів вищої освіти України персональними комп'ютерами навчального призначення у розрахунку на 100 студентів денної форми навчання (Значення 2). Динамічна зміна обох значень позначена протягом періоду з 1991 року до 2011 року.

Рік	Значення	
	1	2
1991	7,83	54,53
1992	8,60	53,97
1993	9,38	32,02
1994	9,19	34,18
1995	8,31	33,81
1996	8,45	32,34
1997	10,81	30,05
1998	12,10	30,04
1999	14,29	29,79
2000	16,90	25,73
2001	20,64	23,09
2002	20,45	22,70
2003	22,57	21,80
2004	23,86	20,93
2005	25,34	19,45
2006	26,16	18,69
2007	26,71	16,12
2008	28,56	13,99
2009	28,98	12,88
2010	29,14	11,16
2011	30,25	11,11

Гістограма динаміки обох часових рядів виглядає так:



Точкова діаграма розсіювання значень часових рядів має такий вигляд:



Визначивши абсолютні прирости (спади) динамічних рядів, зобразимо результат у вигляді гістограми:



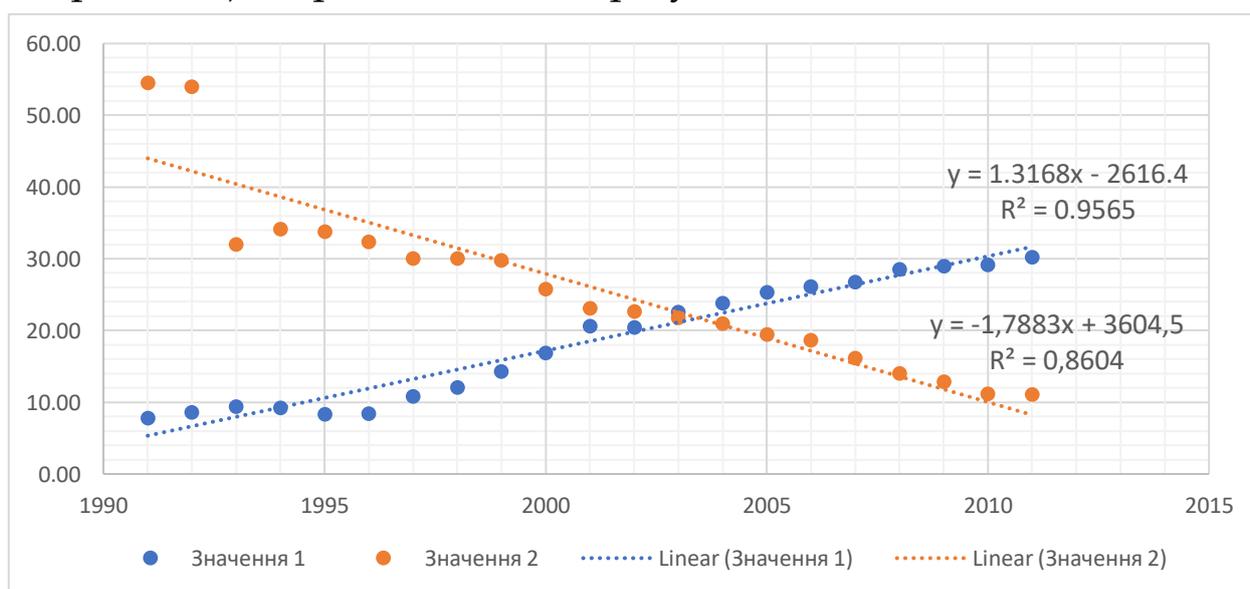
Результати визначення темпів зростання (спадання) часових рядів (у відсотках) можна зобразити у вигляді графіка:



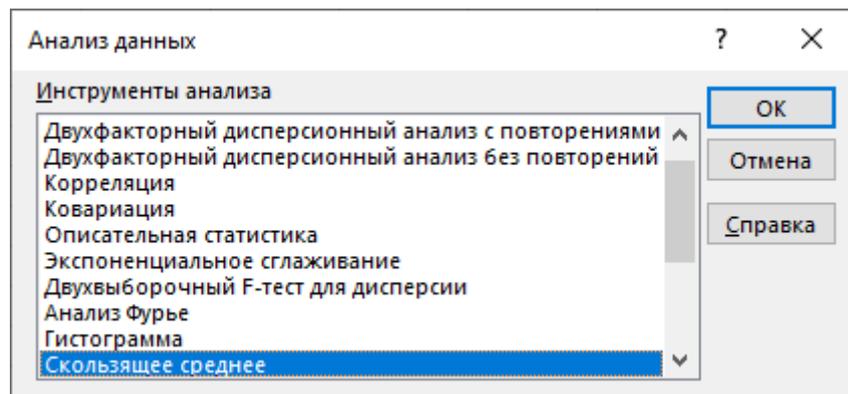
Для наочного зображення темпів приросту (спаду) часових рядів (у відсотках) знову скористаємося гістограмою:



Розглянемо деякі способи вирівнювання даних динамічних рядів. Використавши вирівнювання методом найменших квадратів, маємо змогу зобразити тренди обох динамічних рядів. Зробимо це з використанням точкової діаграми динаміки зміни обох часових рядів. В MSEXCEL тренди зображуються дуже просто: треба на точковій діаграмі клацнути правою клавішею мишки на будь-якому значенні обраного часового ряду, у меню обрати «додати лінію тренду», а в додатковому вікні поставити прапорці навпроти «лінійна», «показати рівняння на діаграмі», «розмістити на діаграмі величину достовірності апроксимації R^2 » (фактично, коефіцієнт детермінації). Отримаємо такий результат:



Отримати ряд, згладжений методом ковзного середнього можна за допомогою надбудови MS EXCEL Пакет аналізу (Analysis ToolPak). Надбудова доступна із вкладки *Дані*, група *Аналіз*. Викличемо надбудову *Пакет аналізу*, оберемо інструмент *Ковзне середнє*.



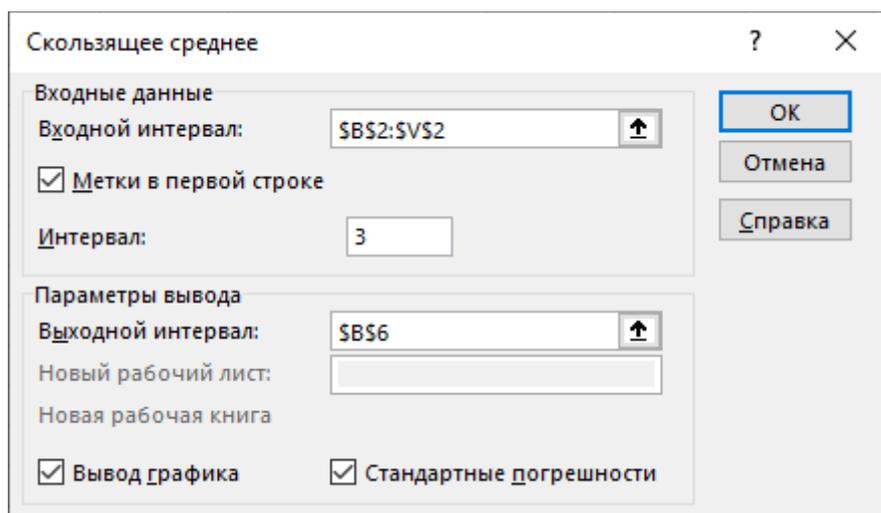
У діалоговому вікні, що з'явилося, в полі *Вхідний інтервал* введемо посилання на діапазон з даними ряду – у нашому випадку для ряду *Значення 2*.

Якщо діапазон включає і заголовок, потрібно встановити галочку в полі *Мітки* в першому рядку. У нашому випадку встановлювати галочку не потрібно, так як заголовок не входить до заявленого діапазону комірок.

У полі *Інтервал* встановимо значення 3 – усереднене значення ряду за 3 періоди. У полі *Вихідний інтервал* достатньо ввести посилання на ліву верхню комірку діапазону з результатами (зазначимо комірку $B\$6$).

Також поставимо галочки у поле *Виведення графіка* та *Стандартні похибки* (буде виведений стовпець з розрахунками похибок, англ. Standard Errors).

В результаті вікно *Ковзне середнє* матиме такий вигляд:



В результаті роботи надбудови, MS EXCEL розмістив значення ряду, отриманого методом ковзного середнього в середньому стовпці (див. таблицю нижче). Лівий стовпчик – це значення ряду *Значення 2*.

У двох верхніх комірках середнього стовпчика містяться текстові значення помилки #Н/Д, так як нами був обраний варіант усереднення за 3 періоди (для отримання першого значення ковзного середнього потрібно 2 попередні та 1 поточне значення). Перше значення середнього стовпчика міститься в третій зверху комірці. Друге значення – в четвертій і т.д.

У правому стовпці, починаючи з п'ятої зверху комірки, MS EXCEL розмістив формули для обчислення похибок (англ. Standard Errors) – у нашому випадку $=\text{КОРІНЬ}(\text{СУМКВРАЗН}(D10:D12;E10:E12)/3)$.

Значення

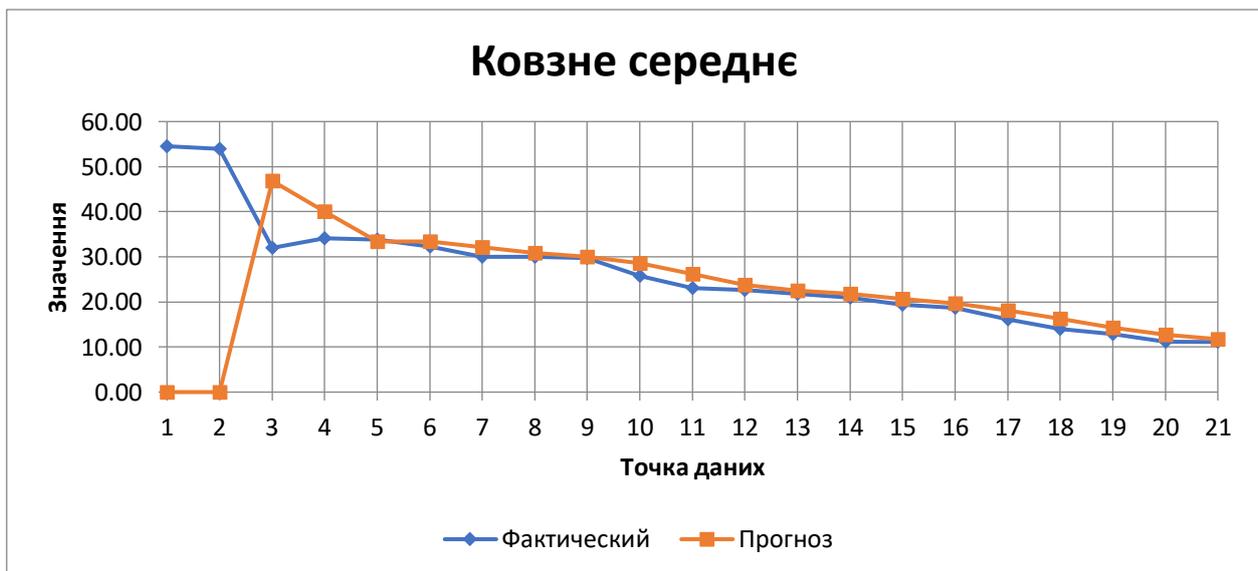
2

54,53	#Н/Д	#Н/Д
53,97	#Н/Д	#Н/Д
32,02	46,84	#Н/Д
34,18	40,05	#Н/Д
33,81	33,33	9,208314
32,34	33,44	3,463242
30,05	32,07	1,354809
30,04	30,81	1,39941
29,79	29,96	1,250844

25,73	28,52	1,674048
23,09	26,20	2,415633
22,70	23,84	2,501769
21,80	22,53	1,960046
20,93	21,81	0,932184
19,45	20,73	0,989474
18,69	19,69	1,065251
16,12	18,09	1,471695
13,99	16,27	1,830391
12,88	14,33	1,928167
11,16	12,68	1,787548
11,11	11,72	1,261061

Для відтворення рядів MS EXCEL створив діаграму типу графік. Згладжений ряд на діаграмі називається «Прогноз» (ряд червоного кольору), хоча він, за великим рахунком, прогнозом не є.

Перші 2 значення згладженого ряду, які дорівнюють помилці #Н/Д, відображаються як 0, і можуть ввести в оману (особливо, якщо наступні значення ряду близькі до 0). Тому їх краще видалити у стовпці.

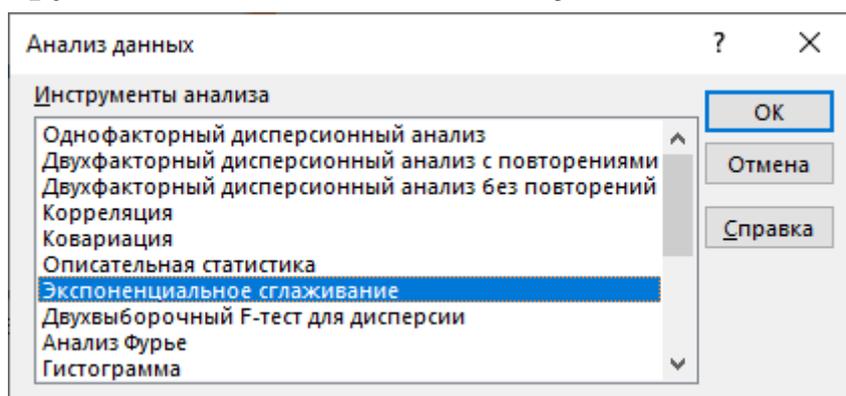


Діаграма дозволяє візуально визначити викиди, тобто. значення вихідного ряду, які суттєво відрізняються від середніх значень. Такі «викиди» можуть бути наслідком помилки, але вони істотно впливають на вигляд згладженого ряду.

Експонентне згладжування використовується для згладжування короткострокових коливань у часових рядах, щоб полегшити визначення довгострокового тренду, а також для прогнозування.

Експонентне згладжування – один з найбільш поширених методів для згладжування часових рядів. На відміну від методу ковзного середнього, де минулі спостереження мають однакову вагу, експоненційне згладжування привласнює їм експоненційно спадні ваги, у міру того як спостереження стають старшими. Іншими словами, останні спостереження дають відносно більшу вагу при прогнозуванні, ніж старі спостереження.

Отримати ряд, згладжений методом експоненційного згладжування можна за допомогою надбудови MS EXCEL Пакет аналізу (Analysis ToolPak). Викличемо надбудову *Пакет аналізу*, оберемо інструмент *Експоненційне згладжування*.



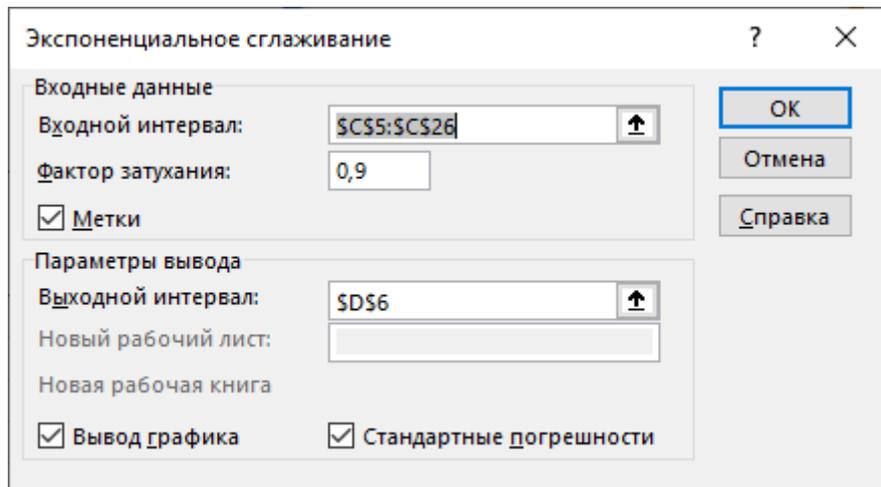
У діалоговому вікні, що з'явилося, в полі *Вхідний інтервал* введемо посилання на діапазон з даними ряду – у нашому випадку для ряду *Значення 2*. Якщо діапазон включає і заголовок, потрібно встановити галочку в полі *Мітки*.

Поле *Фактор згасання* визначає ступінь згладжування ряду. Фактор згасання дорівнює $(1-\alpha)$. Чим більший фактор загасання тим більше згладженим виходить ряд. Встановимо значення 0,9.

У полі *Вихідний інтервал* достатньо ввести посилання на ліву верхню комірку діапазону з результатами.

Також поставимо галочки у поле *Виведення графіка* та *Стандартні похибки* (буде виведений стовпець з розрахунками похибок, англ. Standard Errors).

В результаті вікно *Експоненційне згладжування* матиме вигляд:



В результаті роботи надбудови, MS EXCEL розмістив значення ряду, отриманого методом *Експонентного згладжування*, в середньому стовпці (див. таблицю нижче).

У верхній комірці середнього стовпця міститься текстове значення помилки #Н/Д, так як для отримання першого значення експонентного згладжування потрібно значення вихідного ряду за попередній період.

Перше значення згладженого ряду міститься в другій зверху комірці середнього ряду.

Фактор загасання (0,8) визначає вагу (вклад) попереднього значення згладженого ряду. Відповідно, (1-Фактор згасання)=альфа визначає вагу попереднього значення вихідного ряду.

У правому стовпці, починаючи з 5-ї зверху комірки, MS EXCEL розмістив формули для обчислення похибок (англ. Standard Errors).

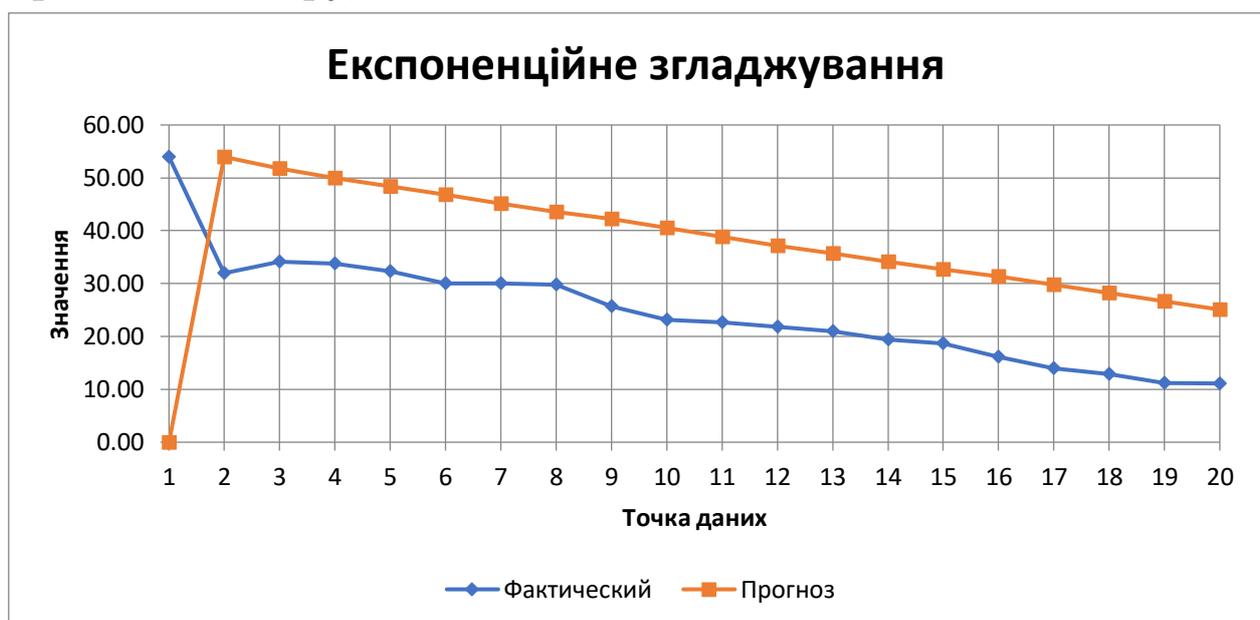
Значення

2

54,53	#Н/Д	#Н/Д
53,97	54,53	#Н/Д
32,02	54,4735196	#Н/Д
34,18	52,2280554	#Н/Д

33,81	50,4228129	16,63751
32,34	48,7615316	19,20145
30,05	47,1196521	17,0436
30,04	45,4126869	16,70265
29,79	43,8754182	16,30204
25,73	42,4668764	15,55733
23,09	40,7931888	15,43634
22,70	39,0228699	16,24731
21,80	37,3905829	16,93086
20,93	35,8315246	16,56206
19,45	34,3413721	15,61578
18,69	32,8522349	15,13137
16,12	31,4360114	14,6558
13,99	29,9044103	14,79755
12,88	28,3129693	15,14835
11,16	26,7696723	15,55662
11,11	25,2087051	15,65361

Для відтворення рядів MS EXCEL створив діаграму типу графік. Згладжений ряд на діаграмі називається «Прогноз» (ряд червоного кольору).



Перше значення згладженого ряду, яке дорівнює помилці #Н/Д, відображається як 0. Тому його краще видалити з комірки.

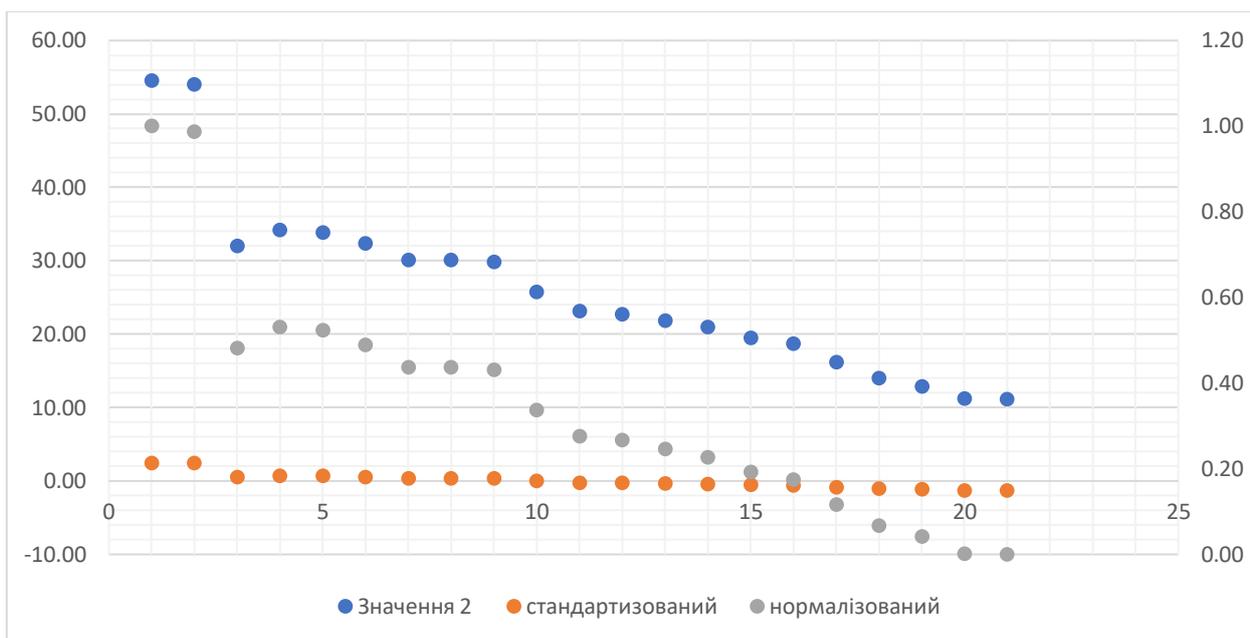
Діаграма дозволяє візуально визначити викиди, тобто значення вихідного ряду, які суттєво відрізняються від середніх значень. Такі «викиди» можуть бути наслідком помилки, але вони істотно впливають на вигляд згладженого ряду.

Проведемо стандартизацію та нормалізацію методом мінімаксного масштабування ряду «Значення 2». Отримаємо такий результат:

Значення 2 стандартизований нормалізований

54,53	2,44	1,00
53,97	2,39	0,99
32,02	0,51	0,48
34,18	0,69	0,53
33,81	0,66	0,52
32,34	0,53	0,49
30,05	0,34	0,44
30,04	0,34	0,44
29,79	0,32	0,43
25,73	-0,03	0,34
23,09	-0,26	0,28
22,70	-0,29	0,27
21,80	-0,37	0,25
20,93	-0,44	0,23
19,45	-0,57	0,19
18,69	-0,64	0,17
16,12	-0,86	0,12
13,99	-1,04	0,07
12,88	-1,13	0,04
11,16	-1,28	0,00
11,11	-1,29	0,00

У вигляді точкової діаграми цей результат виглядає так:

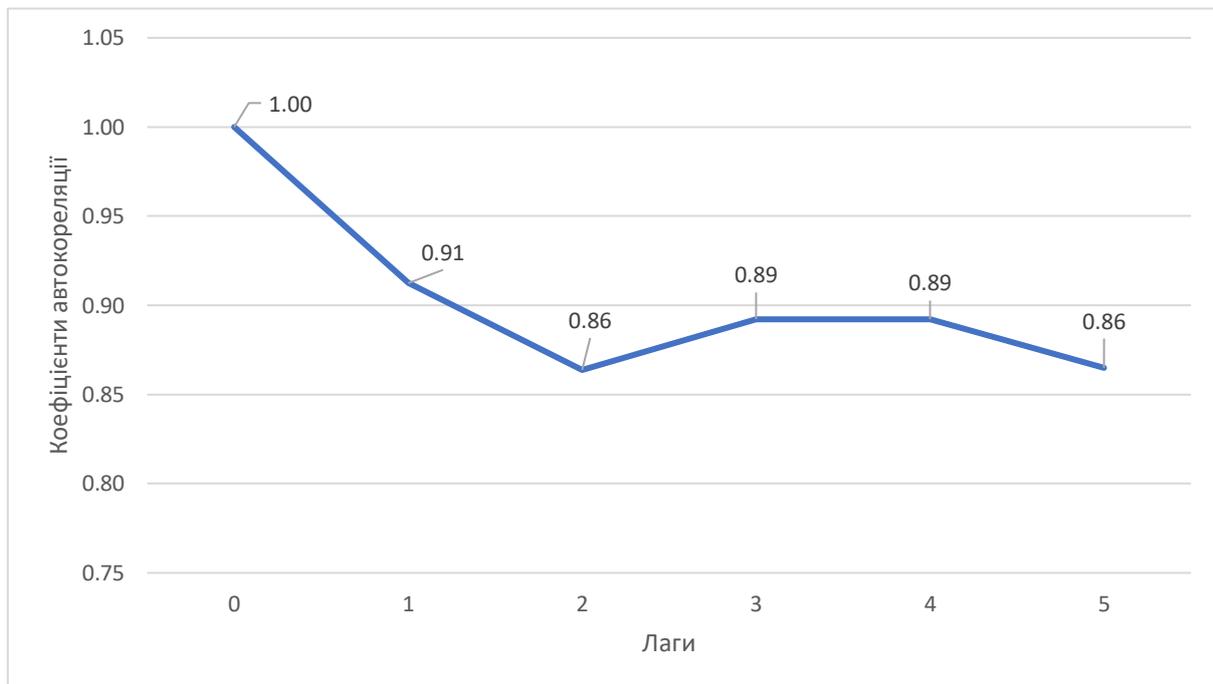


Підрахунок автокореляції ряду «Значення 2» проводимо за допомогою лінійного коефіцієнта кореляції між рівнями вихідного часового ряду та рівнями цього ряду, зрушеними на кілька кроків у часі:

Значення 2	Крок 1	Крок 2	Крок 3	Крок 4	Крок 5
54,53					
53,97	54,53				
32,02	53,97	54,53			
34,18	32,02	53,97	54,53		
33,81	34,18	32,02	53,97	54,53	
32,34	33,81	34,18	32,02	53,97	54,53
30,05	32,34	33,81	34,18	32,02	53,97
30,04	30,05	32,34	33,81	34,18	32,02
29,79	30,04	30,05	32,34	33,81	34,18
25,73	29,79	30,04	30,05	32,34	33,81
23,09	25,73	29,79	30,04	30,05	32,34
22,70	23,09	25,73	29,79	30,04	30,05
21,80	22,70	23,09	25,73	29,79	30,04
20,93	21,80	22,70	23,09	25,73	29,79
19,45	20,93	21,80	22,70	23,09	25,73
18,69	19,45	20,93	21,80	22,70	23,09

	16,12	18,69	19,45	20,93	21,80	22,70
	13,99	16,12	18,69	19,45	20,93	21,80
	12,88	13,99	16,12	18,69	19,45	20,93
	11,16	12,88	13,99	16,12	18,69	19,45
	11,11	11,16	12,88	13,99	16,12	18,69
лаги	0	1	2	3	4	5
кореляція	1,00	0,91	0,86	0,89	0,89	0,86

Значення автокореляційної функції часового ряду «Значення 2» зображені на корелограмі:



Зображена корелограма має виражений максимум при $L=1$, отже ряд містить виражену тенденцію. Це може говорити про те, що ряд містить лінійну тенденцію, яка може бути виражена лінійним коефіцієнтом кореляції.

РОЗДІЛ 6.

ВИМІРЮВАННЯ СОЦІАЛЬНОЇ НЕРІВНОСТІ

6.1. Вимірювання в демографічній та соціальній сферах. Проблема вимірювання соціальної нерівності. Прості та складні гендерні індекси

На сучасному етапі гендерний підхід у теорії та практиці перебуває на стадії розвитку. Основна його концептуальна ідея полягає в тому, що гендер представляється соціальним конструктом, котрий виникає, існує та змінюється під впливом культури та традицій суспільства та адекватно відображає ідею соціального конструювання розходжень між жінками та чоловіками. Саме реалізація гендерного підходу зумовила перегляд існуючої методології в соціальних та соціогуманітарних науках та, відповідно, застосування нових методів. В результаті цього в останні роки набула поширення тенденція формування гендерних сучасних знань з використанням комплексного підходу (гендерний мейнстрімінг) до проблеми рівності жінок та чоловіків. Концепція гендерного мейнстрімінгу стала міжнародно визнаною (Пекінська платформа 1995 року, концепція гендерного мейнстрімінгу Економічної та соціальної ради Генеральної асамблеї ООН 1997 року, концептуальні основи гендерного мейнстрімінгу Ради Європи 1998 року) та покладеною в основу української гендерної політики (Державна програма з утвердження гендерної рівності 2006 року). При цьому деякі дослідники відмічають існування

значних протиріч між сучасною гендерною теорією та концептом гендерного мейнстримінгу, що може негативно вплинути на вироблення механізмів практичної реалізації гендерної рівності в суспільстві [20, с. 358]. Крім комплексного підходу в гендерних дослідженнях використовуються інші методи, вибір яких залежить від заданої мети та предмету дослідження.

Одним з таких методів є теоретичний метод, який передбачає впорядкування знань про гендер на принципах логічної побудови цілісної концепції та визначення місця цих знань як в системі суспільствознавчих наук, так і в рамках окремих соціальних або гуманітарних наук. Теоретичний метод у гендерних дослідженнях дає змогу означити коло наявних проблем та уточнити предмет вивчення, визначити належне місце гендерних знань у сучасних гуманітарних науках у цілому. Історико-генетичний метод щодо гендеру характеризується як гендерний вимір історичного процесу, двоєдиний підхід до оцінювання історичних подій, ситуацій та діяльності в певний період життєдіяльності соціумів та її результатів. Цей метод спрямований на розкриття подій і проблем з позиції реалізації інтересів та потреб обох статей – і жінок, і чоловіків. Соціально-психологічні методи вивчення гендеру передбачають вивчення соціально конструйованих жіночих і чоловічих ролей та відносин, ідентичності та статевих особливостей, психологічних характеристик, тощо. Емпіричний метод дослідження гендеру засвідчує існування й значення факту, реальних подій, конкретних дій, наявність прояву кількісних і якісних параметрів при аналізі гендерних ситуацій. Прогностичний метод при дослідженні гендеру зумовлений необхідністю давати відповіді на питання, що визначило б перспективу гендерного розвитку з огляду на гендерну ситуацію в світі (чи окремому його регіоні) в цілому.

Перелічені методи гендерних досліджень тісно пов'язані з статистичними методами, які передбачають уведення й застосування чисельних показників, що відбивають ступінь справедливості розподілу та користування національними благами

й послугами між соціальними групами в системі загальних досягнень світу або окремого його регіону з урахуванням гендерного чинника. Гендерна статистика вказує на соціально-демографічні особливості таких специфічних груп, як жінки та чоловіки. Показники гендерної статистики є інформаційним джерелом для забезпечення моніторингу становища жінок і чоловіків в світі (чи окремому його регіоні); вони допомагають забезпечити вироблення стратегічних рішень соціально-економічної політики для реалізації принципу рівноправності в суспільстві. Найпростішими показниками гендерної статистики є показники гендерного паритету/асиметрії, які досить широко використовуються в обчисленні більш складних індикаторів. У зв'язку з тим, що показники гендерного паритету/асиметрії дають різні результати при застосуванні їх до генеральних сукупностей та вибірок, виникає потреба вироблення окремих методологічних підходів використання таких показників у демографічній та соціальній статистиці.

Демографічна статистика (використовується й інша назва – статистика населення) – галузь демографії, предметом якої є вивчення статистичних закономірностей відтворення населення. Предметом демографічної статистики є розробка методів статистичного спостереження та вимірювання демографічних явищ і процесів, збирання та обробка статистичної інформації про відтворення населення. *Соціальна статистика* – це галузь статистичної науки, яка вивчає кількісну сторону масових соціальних явищ і процесів у нерозривному зв'язку з їхньою якісною визначеністю з метою виявлення характерних закономірностей, взаємозв'язків і тенденцій. Соціальна статистика вивчає державний устрій, політичну та економічну системи країни, соціальну структуру суспільства, рівень життя населення, споживання матеріальних благ і послуг, умови і характер праці, соціальне забезпечення, торгове, побутове, житлово-комунальне і транспортне обслуговування населення, охорону здоров'я, освіту, культуру і мистецтво, фізичну культуру і спорт [27].

У демографічній статистиці для аналізу складу населення світу, регіонів світу, країн світу чи менших територіальних одиниць застосовується гендерне співвідношення (gender ratio):

$$GR = \frac{population_m}{population_f} \cdot 100$$

яке показує, скільки чоловіків на кожні 100 жінок є у даному регіоні. Надалі під $population_m$ будемо розуміти кількість представників чоловічої статі у відповідному регіоні, а $population_f$ – відповідну кількість представників жіночої статі.

Гендерні співвідношення можуть відрізнятися у різних підгрупах населення. Гендерні співвідношення новонароджених – це зовсім інший показник і відрізняється від гендерного співвідношення для всього населення і від гендерного співвідношення для старшого віку. Біологічно хлопчиків народжується більше ніж дівчат (від 104 до 107 хлопчиків на кожні 100 дівчат). У середньому, жінки більш стійкі до захворювань і живуть довше, ніж чоловіки. Також схильність чоловіків до більш ризикових занять і агресивної поведінки підвищують їх шанси на передчасну смерть. Тому гендерне співвідношення новонароджених показує перевищення кількості чоловіків, а гендерне співвідношення для літніх людей свідчить про перевищення кількості жінок (співвідношення становить менше 100).

У гендерно-нейтральних суспільствах, де чоловіки й жінки живуть в однакових умовах, гендерне співвідношення складає від 98 до 100. У 2015 році, гендерне співвідношення для всього світу складало 102, але по країнах було розсіювання від 274 у Арабських Еміратах до 85 у Латвії та Литві (86 в Україні).

Для аналізу гендерних співвідношень застосовують такі показники й методи: індекс гендерного паритету, індекс гендерного розриву, індекс гендерної асиметрії. Наведемо методику розрахунку цих індексів та проілюструємо особливості їх використання.

Індекс гендерного паритету (gender parity index – GPI) визначається як співвідношення кількості представників жінок у певному територіальному регіоні до кількості чоловіків:

$$GPI = \frac{population_f}{population_m}. \quad (6.1)$$

Цей коефіцієнт використовується як загальноприйнятий коефіцієнт аналізу гендерного паритету, який був введений організацією ЮНЕСКО [5]. Теоретично це співвідношення може змінюватися від 0 (в разі відсутності жінок у групі) до нескінченності (в разі відсутності у групі чоловіків). У разі повної рівності жінок та чоловіків співвідношення рівне 1.

Абсолютний гендерний розрив (absolute gender gap – AGG) вимірюється як різниця між кількістю чоловіків та жінок певному територіальному регіоні:

$$AGG = population_m - population_f. \quad (6.2)$$

Диспропорції на користь жінок позначаються від'ємним значенням гендерного розриву. Гендерний розрив фіксує кількість осіб жінок/чоловіків, якої не вистачає до абсолютного паритету у відповідному регіоні.

Коефіцієнт гендерної асиметрії (gender asymmetry – GA) визначається таким співвідношенням

$$GA = \frac{population_f - population_m}{population_f + population_m}. \quad (6.3)$$

цей коефіцієнт змінюється у проміжку $[-1; 1]$ та вказує на симетрію між чисельністю осіб у чоловічих та жіночих групах у випадку $GA = 0$, на асиметрію в бік жіночої групи у випадку $GA > 0$ (чим ближче значення GA до 1, тим більший дисбаланс у бік представників жіночої статі) та на асиметрію в бік чоловічої групи при $GA < 0$ (при значеннях GA близьких до -1 наявний дисбаланс у бік осіб чоловічої статі).

Розглянемо демографічну ситуацію у ряді країн світу та проаналізуємо прості гендерні показники, а саме: GR, GPI, AGG та GA. Демографічні показники у таблиці 6.1 отримано з [25].

Таблиця 6.1.

Застосування гендерних показників при аналізі демографічної ситуації в окремих країнах

	Country Name	Country Code	Female	Male	GR	GPI	AGG	GA
1	2	3	4	5	6	7	8	9
1	United Arab Emirates	ARE	2524948	6744664	267	0,37	4219716	-0,46
2	India	IND	637879447	686291907	108	0,93	48412460	-0,04
3	China	CHN	668291578	710373422	106	0,94	42081844	-0,03
4	Canada	CAN	18285148	18001277	98	1,02	-283871	0,01
5	United States	USA	163233094	159894419	98	1,02	-3338675	0,01
6	Tunisia	TUN	5769883	5633365	98	1,02	-136518	0,01
7	Greece	GRC	5456150	5290590	97	1,03	-165560	0,02
8	Japan	JPN	64962126	62032385	95	1,05	-2929741	0,02
9	Moldova	MDA	1846518	1705482	92	1,08	-141036	0,04
10	Ukraine	UKR	24201697	20802948	86	1,16	-3398749	0,08

У таблиці країни відсортовані за значенням показника гендерного співвідношення. Всі чотири показники (стовпчики 6–9), які наведені у таблиці, ілюструють різними способами одну й ту ж демографічну ситуацію.

Найбільш зручним показником для візуальної ілюстрації є GA, так як він фіксує симетричні відхилення від нуля в бік 1 у випадку домінування жінок, і в бік -1 у випадку домінування чоловіків.

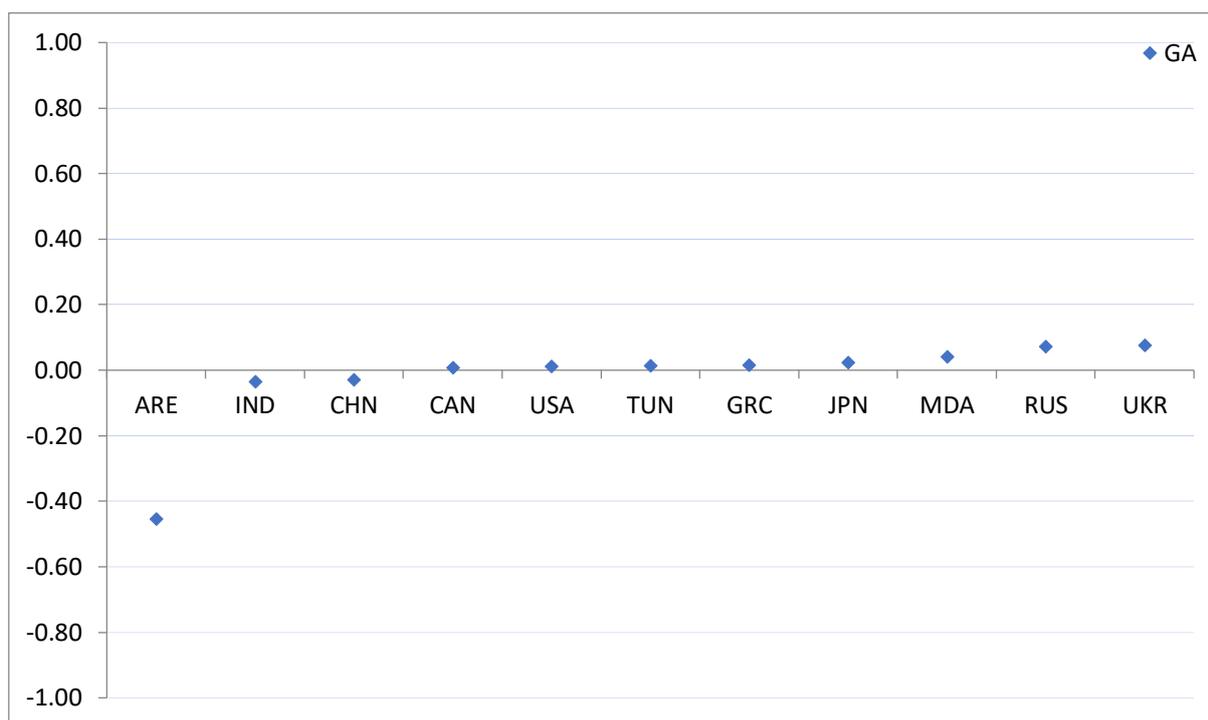


Рисунок 6.1. Візуалізація коефіцієнта гендерної асиметрії
(джерело даних – таблиця 6.1).

У перших трьох країнах (див. таблицю 6.1), бачимо домінування чоловіків у складі населення. Індекс GPI ілюструє те саме, так як він менший одиниці. AGG показує, на скільки чоловіків у країні більше. Від'ємний коефіцієнт асиметрії також говорить про домінування чоловіків.

Середній блок країн (Canada, UnitedStates, Tunisia, Greece, Japan) найбільш близький до паритету у демографічних показниках країни, але є невелике домінування представників жіночої статі.

І останній блок пострадянських країн (Moldova, Ukraine) характеризується високою гендерною асиметрією у бік представників жіночої статі.

Розглянемо ряд індексів, за допомогою яких можна оцінювати питомий гендерний паритет у недемографічних сферах людської діяльності.

В основу обрахунку питомих показників гендерного паритету покладемо відносні величини структури (ВВС), які характеризують

питому вагу складових елементів у загальному підсумку. Їх визначають за формулою:

$$\text{ВВС} = \frac{\text{частина сукупності}}{\text{вся сукупність об'єктів}}$$

Сукупність відносних величин структури показує будову явища, що вивчається. Введемо такі позначення, які ґрунтуються на ВВС:

$$\text{ratio}_f = \frac{\text{number_sph}_f}{\text{population}_f}; \quad \text{ratio}_m = \frac{\text{number_sph}_m}{\text{population}_m} \quad (6.4);$$

де *number_sph* – кількість представників жіночої (чоловічої) статі, які володіють певною ознакою у обраній для аналізу соціальної сфері у певному регіоні, *population* – кількість населення жінок (чоловіків) у даному регіоні.

Індекс питомого гендерного паритету GPI_{sph} визначається таким чином:

$$GPI_{sph} = \frac{\text{ratio}_f}{\text{ratio}_m} \quad (6.5)$$

Цей індекс характеризує, у скільки разів частка жінок, що володіє певною ознакою у обраній для аналізу соціальної сфері, більша/менша від відповідної частки чоловіків. Таким чином, GPI_{sph} змінюється у межах $[0; +\infty]$ та є якісною мірою відношення розподілу характерних ознак у гендерних групах. Випадки, коли $GPI_{sph} = 1$, будемо називати паритетом, а коли GPI_{sph} відмінний від 1 – відсутністю паритету (причому, якщо $GPI_{sph} > 1$, то з переважанням частки жінок, в іншому випадку – з переважанням частки чоловіків).

Абсолютний питомий гендерний розрив AGG_{sph} вимірюється як різниця частки чоловіків від їх загального числа у відповідній демографічній групі та частки жінок (у %):

$$AGG_{sph} = (\text{ratio}_m - \text{ratio}_f) \cdot 100\% \quad (6.6)$$

Цей індекс має зміст, якщо кількість чоловіків та жінок у відповідних демографічних групах приблизно однакова

($population_m \approx population_f$). Індекс характеризує, на скільки відсотків частка чоловіків, що володіє певною характерною ознакою, більша/менша від відповідної частки жінок.

Коефіцієнт питомої гендерної асиметрії GA_{sph} визначається за часткою жінок від їх числа у відповідній демографічній групі $ratio_f$ та часткою чоловіків $ratio_m$ за заданим полем аналізу, причому:

$$GA_{sph} = \frac{ratio_f - ratio_m}{ratio_f + ratio_m} \quad (6.7)$$

Таким чином, GA_{sph} змінюється у межах $[-1; 1]$ та також є якісною мірою відношення розподілу характерних ознак у гендерних групах. Випадки, коли $GA_{sph} = 0$, будемо називати симетрією між частками чоловіків та жінок, а коли він відмінний від 0 – асиметрією (причому, якщо $GA_{sph} > 0$, то з переважанням частки жінок, в протилежному випадку – з переважанням частки чоловіків).

Примітка. Здійснивши елементарні перетворення, встановимо зв'язок між індексом питомого гендерного паритету GPI_{sph} та індексом питомої гендерної асиметрії GA_{sph}

$$(1 + GPI_{sph})(1 - GA_{sph}) = 2.$$

З'ясуємо, 1) наскільки універсальним є застосування індексів GPI_{sph} , AGG_{sph} та GA_{sph} , а також індексів GPI , AGG або GA ; 2) наскільки широкою чи специфічною є сфера їх застосувань; 3) наскільки коректні ми отримаємо при цьому результати.

Продемонструємо використання показників GPI_{sph} , AGG_{sph} , GA_{sph} та показників GPI , AGG , GA на масивах даних різного обсягу при аналізі показників довільної сфери. Перший приклад буде пов'язаний з визначенням співвідношень між жіночими та чоловічими групами, що є учнями початкової школи (Enrolment in primary education) в різних регіонах нашої планети. Обсяги масивів даних будуть обмежуватися такими територіями: Кіровоградська область України, вся Україна, Європа та Північна Америка (UNESCO Region), увесь світ. Чисельні значення кількості осіб, що

були 2012 року учнями початкової школи у кожному з вказаних регіонів, розбивка вказаних осіб за гендерним принципом, загальні демографічні показники та розподіли населення за статтю Кіровоградської області України, всієї України, Європи з Північною Америкою та всього світу станом на 2012 рік представлені у таблиці 6.2. Перелічені регіони у математичному розумінні поняття множини є послідовними підмножинами (або пов'язані співвідношенням включення, якщо рухатися від регіону з найбільшим масивом даних до регіону з найменшим масивом). Таким чином, демографічний масив даних (так само, і масив даних, пов'язаний з навчанням у початковій школі), що складає населення Кіровоградської області, є підмножиною масиву даних, що складає населення України, масив даних населення України включається до масиву даних Європи спільно з Північною Америкою, останній масив, у свою чергу, включається до масиву даних усього світу.

Таблиця 6.2.

Застосування гендерних показників на вибірках різної величини
(на прикладі масиву даних учнів початкової школи
(*Enrolment in primary education, number*))

Дані за 2012 рік	Кіровоградська область	Україна	Європа + Північна Америка	Увесь світ
Дітей у початковій освіті				
$number_{sph}_f$	16 932	775 316	35 793 423	339 733 393
$number_{sph}_m$	17 456	809 066	40 400 077	371 557 319
Загальна демографія				
$population_f$	539 900	24 549 634	620 365 541	3 516 939 678
$population_m$	456 100	21 043 666	588 555 526	3 580 460 987
Прості гендерні показники				
GPI	0,97	0,96	0,89	0,91
AGG	524	33 750	4 606 654	31 823 926
GA	-0,02	-0,02	-0,06	-0,04
GPI_{sph}	0,82	0,82	0,84	0,93
AGG_{sph}	0,0069	0,0069	0,0109	0,0072
GA_{sph}	-0,10	-0,10	-0,09	-0,04

У даному прикладі продемонстровано некоректне застосування індексів GPI , AGG та GA до показників соціальної сфери, тобто розрахунок індексів паритету для освітнього показника був зведений до таких спрощених формул:

$$GPI = \frac{number_{sph_f}}{number_{sph_m}}; \quad (6.8)$$

$$AGG = number_{sph_m} - number_{sph_f}; \quad (6.9)$$

$$GA = \frac{number_{sph_f} - number_{sph_m}}{number_{sph_f} + number_{sph_m}}. \quad (6.10)$$

З даних простих гендерних показників таблиці 2 бачимо, що на великому масиві даних (увесь світ) коефіцієнти GA та GPI_{sph} працюють однаково (рівні $-0,04$), а індекси GPI та GPI_{sph} є досить близькими (мають значення, відповідно, $0,91$ та $0,93$) і знаходяться в межах похибки обчислень. Цього не спостерігається на масивах меншої величини. По Кіровоградській області та по Україні різниця в показниках складає: для GPI та GPI_{sph} $0,14-0,15$, а для GA та GPI_{sph} $0,08$. Дещо менша різниця в показниках для Європи разом з Північною Америкою – вона складає відповідно $0,05$ та $0,03$. Тобто для визначення показників гендерного паритету та гендерної асиметрії серед учнів початкової школи в усьому світі можна користуватися як формулами (6.5) та (6.7), так і спрощеними формулами (6.8) та (6.10) (тобто, лише у випадках, коли $population_m \approx population_f$). А для визначення таких же показників серед учнів початкової школи в інших регіонах (з меншим масивом даних) доцільно використовувати формули (6.5) та (6.7), так як вони дають результат, який відповідає демографічній ситуації у відповідному територіальному регіоні. В іншому випадку ми отримуємо результат, в якому повністю ігноруються демографічні гендерні співвідношення.

Розглянемо інший приклад, який буде пов'язаний з визначенням співвідношень між жіночими та чоловічими групами, що станом на 2016 рік представляли економічно активне населення

серед демографічної групи населення, якому більше, ніж 15 років. Обсяги масивів будуть обмежуватися такими ж територіями: Кіровоградська область України, вся Україна, Європа та Північна Америка, увесь світ (таблиця 6.3).

Таблиця 6.3.

Застосування гендерних показників на вибірках різної величини (на прикладі масиву даних економічно активного населення 15+ (Labor force participation rate 15+))

Дані за 2016 рік	Кіровоградська область	Україна	Європа + Північна Америка	Увесь світ
Економічно активне населення 15+				
$number_sph_f$	202 700	9 855 331	198 563 366	1 344 090 821
$number_sph_m$	226 100	10 914 912	240 024 669	2 071 653 076
Загальна демографія 15+				
$population_f$	447 189	20 876 809	389 932 647	2 750 154 246
$population_m$	366 546	17 273 031	357 943 440	2 750 152 789
Прості гендерні показники				
GPI	0,90	0,90	0,83	0,65
AGG	23 400	1 059 581	41 461 303	727 562 255
GA	-0,05	-0,05	-0,09	-0,21
GPI_{sph}	0,73	0,75	0,76	0,65
AGG_{sph}	0,1636	0,1598	0,1613	0,2646
GA_{sph}	-0,15	-0,14	-0,14	-0,21

З даних таблиці 6.3 бачимо, що аналогічно до попереднього випадку на великому масиві даних (увесь світ) і індекси GA та GPI_{sph} , і індекси GPI та GPI_{sph} працюють однаково (вони рівні відповідно 0,65 та -0,21). І знову цього не спостерігається на масивах меншої величини. По Кіровоградській області та по Україні різниця в показниках складає: для індексів GPI та GPI_{sph} 0,15–0,17, а для індексів GA та GPI_{sph} 0,09–0,10. Дещо менша різниця в показниках для Європи разом з Північною Америкою – вона складає відповідно 0,07 та 0,05. В цьому випадку також для визначення показників гендерного паритету та гендерної асиметрії серед економічно

активного населення 15+ усього світу можна користуватися як формулами (6.5) та (6.7), так і формулами (6.8) та (6.10) (знову, лише у випадках, коли $population_m \approx population_f$). А для визначення таких же показників серед економічно активного населення 15+ в інших регіонах (з меншим масивом даних) доцільно використовувати формули (6.5) та (6.7).

Вся проблема визначення методики використання тих чи інших формул визначення гендерного паритету чи гендерної асиметрії полягає у оцінці загальних демографічних показників того регіону, який вивчається. З формул (6.5) та (6.7) зрозуміло, що чим ближчі одні до одного загальні демографічні показники чоловіків та жінок ($population_m \approx population_f$), тим ближчим результат роботи формул (6.5) та (6.7) буде до результату роботи формул відповідно (6.8) та (6.10). Проаналізуємо значення співвідношення

$$\frac{population_f - population_m}{population_f + population_m} \quad (6.11)$$

для демографічних показників обох прикладів (таблиця 4).

Таблиця 6.4.

Оцінка відхилення значень $population_f$ та $population_m$ в розрахунку на загальну чисельність населення відповідної групи

	Кіровоградська область	Україна	Європа + Північна Америка	Увесь світ
Загальна демографія (приклад 1)				
$population_f$	539 900	24 549 634	620 365 541	3 516 939 678
$population_m$	456 100	21 043 666	588 555 526	3 580 460 987
Загальна демографія 15+ (приклад 2)				
$population_f$	447 189	20 876 809	389 932 647	2 750 154 246
$population_m$	366 546	17 273 031	357 943 440	2 750 152 789
Значення співвідношення (10)				
Приклад 1	8,4E-2	7,7E-2	2,6E-2	-8,9E-3
Приклад 2	9,9E-2	9,4E-2	4,3E-2	2,65E-7

І в першому, і в другому випадках відхилення значень $population_f$ та $population_m$ (відповідно кількості чоловіків та жінок в

загальній демографії) у регіоні «увесь світ» становить менше 1 % в розрахунку на загальну чисельність населення відповідної групи (в першому випадку це населення всього світу, а в другому – населення світу 15+). В інших випадках (регіонах) відхилення значно більші (від 2-х до 10-ти відсотків).

Таким чином, аналіз результатів застосування простих гендерних індексів дає можливість стверджувати, що аналіз гендерного паритету/асиметрії потребує окремого методологічного підходу у демографічній статистиці і окремого підходу у соціальній статистиці.

По-перше, при обрахунку простих гендерних показників у демографічній статистиці використання формули індексу гендерного паритету GPI (співвідношення (6.1)) та коефіцієнта гендерної асиметрії GA (співвідношення (6.3)) є коректним та загальноновживаним.

По-друге, при обрахунку простих гендерних показників у соціальній статистиці використання формули індексу гендерного паритету GPI (співвідношення (6.1) або (6.8)) та коефіцієнта гендерної асиметрії GA (співвідношення (6.3) або (6.10)) не є коректним та загальноновживаним. При таких обрахунках коректним є використання формули індексу питомого гендерного паритету GPI_{sph} (співвідношення (6.5)) та коефіцієнта питомої гендерної асиметрії GA_{sph} (співвідношення (6.7)), які враховують демографічні показники $population_f$ та $population_m$.

По-третє, в тих випадках, коли виконується співвідношення $population_m \approx population_f$ (необхідну точність обчислень можна задавати – див. співвідношення (6.10) та таблицю 6.4), використання формули індексу гендерного паритету GPI (співвідношення (6.8)) та коефіцієнта гендерної асиметрії GA (співвідношення (6.10)) в соціальній статистиці можна вважати повністю коректним.

Для гендерного аналізу, який охоплює одночасно декілька сфер діяльності людей, застосовують *складні індекси*. В основному ці індекси розраховують на рівні країни і використовують їх для ранжування країн за рівнем людського розвитку. Основними

складними індексами, за якими оцінюється гендерна рівність у країні в цілому, є такі: індекс гендерного розвитку (gender development index), індекс реалізації прав жінок (gender empowerment measure), індекс гендерної нерівності (gender inequality index), індекс гендерної рівності (Gender Equality Index), індекс глобального гендерного розриву (global gender gap index), індекс соціальних інститутів та гендеру (Social Institutions and Gender Index). Методики обрахунку цих індексів математично складні і часто містять у своїх проміжних розрахункових формулах індекс гендерного паритету. Наприклад, для розрахунку індексу глобального гендерного розриву (global gender gap index) [14], спочатку розраховується індекс гендерного паритету за 14-ма базовими показниками країни, а потім на їхній основі розраховується складний індекс.

У більшості міжнародних звітів, які аналізують гендерні відмінності у країнах та регіонах світу, виділяють ряд основних сфер, у яких проводять гендерний аналіз. Наприклад, для розрахунку індексу гендерної рівності (Gender Equality Index) виділяють 6 основних сфер (робота, гроші, знання, час, влада і здоров'я) та 2 допоміжні сфери (перехресні нерівності та насилля) [15]. Індекс глобального гендерного розриву (global gender gap index) аналізують і розраховують за 14-ма індикаторами у 4-х сферах: економічна участь і кар'єрні можливості; освіта; здоров'я; політичні права та можливості [14]. Цей перелік індексів і сфер, які вони охоплюють, можна продовжити. Тобто гендерний статистичний аналіз охоплює практично всі сфери людської діяльності і потребує чіткої методики оцінки гендерного паритету/асиметрії у кожній із цих сфер.

Розглянемо теоретичні особливості застосування математико-статистичних методів для розрахунку *індексу гендерного розвитку (Gender Development Index)*. Індекс гендерного розвитку охоплює аспекти тривалості життя (довголіття), освіти (знання) і гідного рівня життя. Індекс гендерного розвитку (Gender Development Index, GDI) відображає гендерний розрив в людському розвитку

країни за трьома основними показниками: здоров'я, освіта та контроль над економічними ресурсами [13]. При підрахунку GDI за основу береться Індекс людського розвитку (Human Development Index, HDI) і концентрується увага на гендерних розривах. У доповіді експертів ООН «Концепція і вимірювання людського розвитку» за 1990 р. вперше була опублікована методика розрахунку Індексу людського розвитку [7], а 1995 року була опублікована методика розрахунку індексу розвитку людського потенціалу з урахуванням гендерного фактору – GDI [9]. GDI оцінює людський розвиток жінок та чоловіків і водночас враховує гендерну нерівність між цими показниками [13]. Людський розвиток, своєю чергою, розраховується за трьома основними вимірами:

- довголіття, що визначається очікуваною тривалістю життя при народженні;
- знання, що оцінюються за очікуваною та середньою тривалістю шкільного навчання;
- гідний рівень життя, що характеризується наближеним показником отриманого доходу на одну особу.

Індекс гендерної нерівності (GII – Gender Inequality Index), складений Програмою розвитку Організації Об'єднаних Націй, був вперше опублікований у 2013 році і заснований на тій думці, що надто часто жінки та дівчатка зазнають дискримінації в галузі охорони здоров'я, освіти та на ринку праці з негативними наслідками для їхньої свободи (Програма розвитку Організації Об'єднаних Націй (2018 рік). Індекс є сукупним показником гендерної нерівності за трьома вимірами: репродуктивне здоров'я; розширення прав та можливостей жінок та їх участь на ринку праці [8]. Мета GII полягає у кількісній оцінці чи визначенні вартості гендерної рівності з погляду розвитку людського потенціалу. Чим вище значення індексу, тим більша нерівність між жінками та чоловіками і тим більші втрати для людського розвитку.

Індикатори, на які спирається індекс GII, не мають відношення до загального рівня розвитку країни, тому країни, що

розвиваються, можуть показати порівняно непогані результати, якщо в становищі чоловіків і жінок немає великих відмінностей.

Значення GII коливаються від нуля (повна рівність жінок та чоловіків) до одиниці (повна нерівність). Індекс показує відсоток втрат потенційного рівня людського розвитку, зумовлених гендерною нерівністю.

Репродуктивне здоров'я тут вимірюється двома показниками: рівнем материнської смертності й коефіцієнтом народжуваності у підлітковому віці (молодше 20 років). Чим вони нижчі, тим ближче країна до гендерної рівності; за задумом авторів індексу, в ідеальній ситуації ні материнської смертності, ні народжуваності серед підлітків не повинно бути. Материнські смерті – здебільшого запобіжні – є яким сигналом про статус жінок у суспільстві. Небезпека смерті при дитинонародженні знижується завдяки освіті, адекватному харчуванню, доступності контрацепції, допологової медичної кваліфікованої допомоги. Раннє материнство завдає шкоди здоров'ю та обмежує майбутні можливості – заважає молодим жінкам здобувати освіту та часто примушує займатися низькокваліфікованою роботою.

«Розширення прав та можливостей» також вимірюється двома показниками: співвідношенням числа жінок і чоловіків у парламенті та рівнем середньої та вищої освіти серед жінок та чоловіків. У політичній сфері жінки традиційно перебувають у залежному становищі на всіх рівнях управління, але GII враховує лише участь жінок у роботі парламенту. Доведено, що підвищення рівня освіти розширює свободу жінок, зміцнюючи їхню здатність ставити питання, розмірковувати, дієво впливати на умови свого життя. Освічені жінки мають більше шансів отримати роботу, що задовольняє їх, брати участь у громадських дискусіях, піклуватися про своє здоров'я та здоров'я своєї сім'ї та виступати з іншими ініціативами.

Економічна активність вимірюється участю жінок у робочій силі – порівняльними показниками економічної активності жіночого та чоловічого населення.

Індекс гендерного розриву (*Global Gender Gap Index*) – глобальне дослідження та рейтинг країн світу, що його супроводжує, за інтегральним показником рівноправності статей. Розраховується за методикою Всесвітнього економічного форуму (*World Economic Forum*), що ґрунтується на комбінації загальнодоступних статистичних даних у галузі соціально-економічного розвитку по різних країнах світу. Історія існування GGGI бере свій початок із травня 2005 року, коли учасники Світового економічного форуму розпочали своє перше дослідження, намагаючись оцінити розмір гендерного розриву в 58 країнах за допомогою економічних, освітніх, медичних та політичних критеріїв. У результаті дослідження перший звіт про гендерні розриви за 2006 рік містив унікальну методологію побудови GGGI[6].

Індекс призначений для вимірювання гендерних відмінностей у доступі до ресурсів та можливостей в окремих країнах, а не фактичного рівня наявних ресурсів та можливостей у цих країнах. Таким чином, GGGI вимірює лише гендерні відмінності у країнах світу незалежно від рівня їхнього розвитку. Передбачається, що GGGI має використовуватися державами, які прагнуть скорочення гендерної нерівності, як інструмент для аналізу проблемних моментів у їхній соціальній політиці, розробки заходів щодо скорочення гендерного розриву та відстеження їх ефективності з часом.

GGGI вимірює рівень гендерного розриву, який існує в тих чи інших країнах між жінками та чоловіками, по 14 різних змінних у чотирьох ключових областях:

- Економічна участь та кар'єрні можливості (індикатори: рівень участі в робочій силі (%), співвідношення заробітної плати, приблизний зароблений дохід за ПКС (дол. США), законодавці, вищі посадові особи та менеджери (%), професійно-технічні працівники (%)).
- Освіта (індикатори: рівень грамотності (%), зарахування до початкової освіти (%), зарахування до середньої освіти (%), зарахування до вищої освіти (%)).

- Здоров'я та виживання (індикатори: співвідношення статей при народженні, очікувана тривалість здорового життя (роки)).
- Політичні права та можливості (індикатори: жінки в парламенті (%), жінки на міністерських посадах (%), роки з жінкою на чолі держави (останні 50 років)).

Всі розрахунки в GGGI виконуються на основі Індексу Гендерного Паритету (GPI), який розраховується як співвідношення показника кількості жінок до показника кількості чоловіків.

6.2. Застосування інструментів EXCEL для аналізу простих та складних гендерних індексів

При використанні інструментів Excel для аналізу гендерних індексів на першому етапі слід підготувати дані. Підготовка даних для аналізу гендерних показників в Excel є ключовим етапом, що включає їх очищення, впорядкування та форматування. Розглянемо кроки, які допоможуть підготувати дані для ефективного аналізу:

1. *Очищення даних.* З метою очищення даних слід перевірити їх на наявні помилки й порожні клітинки – видалити або виправити пусті або неправильні значення. Далі слід стандартизувати формати даних – перевірити, щоб усі дані були представлені в однакових одиницях (відсотках, цілих числах тощо). Нарешті, слід видалити повторювані записи, що можуть спотворювати аналіз.

2. *Проведення нормалізації та кодування даних.* Кодування категоріальних даних (якщо дані містять категорії (наприклад, «чоловік» і «жінка») слід за допомогою чисел (1 для чоловіків, 2 для жінок) або залишити їх у текстовому форматі залежно від потреб.

Для нормалізації показників, якщо потрібно порівнювати показники різних масштабів (наприклад, зарплата та рівень освіти),

слід привести їх до одного масштабу. Для цього можна використовувати формулу нормалізації. Формула для лінійної нормалізації є такою:

$$\text{Нормалізоване значення} = \frac{X - X_{min}}{X_{max} - X_{min}},$$

де X – значення показника, X_{min} – мінімальне значення показника, X_{max} – максимальне значення показника.

Можуть використовуватися інші способи нормалізації. Наприклад, Z-нормалізація (стандартизація) – цей метод використовують, коли необхідно врахувати відхилення значень від середнього і зробити їх сумісними з нормальним розподілом. Після стандартизації середнє значення показника дорівнює 0, а стандартне відхилення – 1. Формула для Z-нормалізації виглядає так:

$$Z - \text{нормалізоване значення} = \frac{X - \mu}{\sigma},$$

де X – значення показника, μ – середнє значення, σ – стандартне відхилення.

Або нормалізація за відсотками (перцентилі) – у цьому підході значення показників переводяться у відсотки, що дозволяє ранжувати заклади освіти та порівнювати їх за відносною позицією. Відсоткове значення показує, який відсоток закладів знаходиться нижче за певне значення.

Можлива також логарифмічна нормалізація – застосовується для показників з великим розкидом значень або у випадках, коли розподіл є асиметричним. Логарифмічна нормалізація зменшує вплив великих значень і стабілізує дисперсію. Формула для логарифмічної нормалізації є такою:

$$\text{Нормалізоване значення} = \log_{10}(X + 1)$$

Нарешті, можна використати нормалізація за допомогою відносних відстаней – застосовується, коли потрібно оцінити значення показника щодо середнього або максимально можливого.

Наприклад, можна розраховувати відношення показника до середнього по вибірці чи еталонному значенні.

Щодо вибору методу нормалізації зауважимо таке:

- Лінійна нормалізація підходить, коли всі показники повинні мати однакову шкалу для подальшого зваженого сумування.
- Z-нормалізація рекомендується для показників з нормальним розподілом.
- Логарифмічна нормалізація зручна для даних з великим розкидом або для значень, які мають асиметричний розподіл.
- Нормалізація за перцентилями та нормалізація за допомогою відносних відстаней дозволяє аналізувати відносну позицію кожного закладу освіти серед інших.

Обраний метод нормалізації залежить від типу показників, мети аналізу і подальших кроків у розрахунку індексу якості навчання.

3. *Перетворення дат.* Якщо дані містять дати, слід переконатися, що всі вони в одному форматі (наприклад, ДД/ММ/РРРР). В іншому випадку слід перетворити текстові дати у необхідний чи обраний єдиний формат дати.

4. *Використання текстових функцій для форматування.* Іноді записи можуть містити зайві пробіли або зайві символи. Слід використати відповідні функції, щоб вирівняти текст у клітинках.

5. *Впорядкування та структурування даних у таблицю.* Для створення таблиці слід у текстовому редакторі Word виділити діапазон даних і створити таблицю (вкладка «Вставка» → «Таблиця»). Створену таблицю легко перенести до Excel, скопіювавши її повністю та перенісши у відповідний розмір таблиці в Excel. Це дасть можливість сортувати, фільтрувати та автоматично застосовувати формули до нових рядків.

Далі слід додати чіткі заголовки, які мають відображати зміст стовпця (наприклад, «Стать», «Вік», «Зарплата») до кожної колонки створеної таблиці.

6. *Підготовка зведених даних.* Для організації аналізу, розміщених у вихідній таблиці даних та для застосування відповідних формул (меню «Формули»), слід створити нові стовпці з обчислюваними значеннями, якщо потрібно визначити, наприклад, різницю у зарплатах між чоловіками та жінками і т.п..

7. *Створення зведених таблиць для первинного аналізу.* Часто доцільним для аналізу є створення зведених таблиць, щоб швидко групувати й порівнювати дані за статтю або іншими категоріями, що допоможе отримати загальну картину й виявити відмінності ще до більш глибокого аналізу.

8. *Перевірка даних на коректність.* Доцільно використовувати умовне форматування для виявлення аномальних значень (наприклад, у випадках, коли показники заробітної плати нижчі мінімальної або показники віку виходять за межі очікуваного діапазону). Використання умовного форматування для наочного порівняння показників означає застосування спеціального форматування в Excel для того, щоб автоматично виділити певні дані на основі заданих критеріїв. Це дозволяє швидко помітити важливі закономірності, відмінності або аномалії в наборі даних, що може бути корисним при порівнянні гендерних показників.

Як умовне форматування допомагає в порівнянні гендерних показників?

1. Виділення гендерних різниць – наприклад, якщо ви аналізуєте рівень доходів за статтю, можна налаштувати умовне форматування так, щоб значення вищі за певну суму виділялися одним кольором, а нижчі – іншим. Це полегшує ідентифікацію груп із вищими або нижчими доходами.

2. Візуалізація значень через кольорові шкали – кольорові шкали дають змогу створити градацію кольорів, що відповідають рівням показників. Наприклад, можна застосувати червоно-зелений градієнт, де найнижчі значення позначаються червоним, середні – жовтим, а найвищі – зеленим. Це особливо корисно для порівняння, наприклад, індексу гендерної рівності між різними регіонами.

3. Піктограми і стовпчики даних – використання піктограм (стрілок, кружечків тощо) або стовпчиків даних допомагає наочно порівняти значення між категоріями. Наприклад, для відображення різниці в показниках здоров'я чи освіти між чоловіками та жінками можна використати іконки: зелена стрілка вгору для вищих значень, червона стрілка вниз для нижчих.

4. Виділення певних умов (правил) – якщо потрібно виділити певні значення (наприклад, показники, що перевищують середній рівень), умовне форматування дозволяє задати правила, які виділяють ці значення кольором або рамкою. Це дає змогу швидко помітити відмінності, наприклад, якщо певний гендерний показник в окремій групі значно відрізняється від загального середнього.

5. Застосування умов для групи значень – можна застосувати умовне форматування для групи значень (наприклад, рядків або колонок), що допомагає порівняти їх на загальному тлі. Наприклад, у таблиці із гендерними індексами країн можна виділити червоним кольором показники нижче певного порогового значення, що допомагає швидко оцінити країни з найнижчими показниками.

Приклади налаштування умовного форматування в Excel:

А) налаштування кольорових шкал – слід обрати потрібний стовпець, натиснути «Головна» → «Умовне форматування» → «Колірні шкали», і обрати градацію кольорів для наочного порівняння.

Б) правила виділення певних значень – наприклад, слід обрати діапазон даних, перейти до «Умовне форматування» → «Правила виділення клітинок» → «Більше ніж», щоб виділити значення вищі за певний поріг.

В) додавання іконок або виділення стовпчиків даних – слід обрати «Умовне форматування» → «Панелі даних» або «Набори значків», щоб додати піктограми чи смуги, які візуалізують порівняння.

Цей підхід допомагає зосередитися на значеннях, що потребують уваги, і виявити гендерні тенденції, які можуть бути

важкими для сприйняття у вигляді лише чисел. Це також спрощує порівняння великих масивів даних і робить аналіз доступнішим для візуального сприйняття.

Для аналізу простих гендерних індексів використовуються базові показники (середні значення, медіани, стандартні відхилення – закладка «Формули»), а також специфічні формули (6.1)–(6.6). Крім того, для аналізу використовується візуалізація гендерних даних. Візуалізація гендерних даних допомагає швидко побачити відмінності та тенденції між групами. В Excel є кілька підходів і типів діаграм, які ефективно показують гендерні показники.

1. *Стовпчасті діаграми (Column Charts)* підходять для порівняння значень між групами, наприклад, середнього доходу чоловіків та жінок. Для її створення слід виділити дані для порівняння (наприклад, колонки із заробітною платою та статтю), перейти на вкладку «Вставити» → «Стовпчаста діаграма» і вибрати потрібний тип.

2. *Складені стовпчасті діаграми (Stacked Column Charts)* використовуються для демонстрації співвідношення всередині груп (наприклад, частка чоловіків та жінок у різних вікових групах). Для їх створення слід обрати «Зібраний стовпець» у розділі стовпчастих діаграм.

3. *Секторні діаграми (Pie Charts)* застосовують для відображення частки або відсоткового співвідношення, наприклад, частки чоловіків та жінок у компанії або організації. Для їх створення слід виділити дані й перейти до «Вставити» → «Секторна діаграма». Excel автоматично створить діаграму, що показує відсотковий розподіл.

4. *Гістограми (Histograms)* дозволяють показати розподіл даних для кожної статі, наприклад, розподіл доходів серед чоловіків і жінок. Для створення слід у вкладці «Вставити» обрати «Гістограма». Далі обрати діапазон даних для побудови гістограми за певним показником для чоловіків і жінок окремо.

5. *Лінійні діаграми (Line Charts)* використовують для демонстрації змін у показниках з часом для обох статей, наприклад, зміну рівня зайнятості чоловіків і жінок. Для створення слід виділити дані, натиснути «Вставити» виберіть «Гістограма». Далі слід ввести дані по осях (вік або час на горизонтальній осі, показник – на вертикальній).

6. *Точкові діаграми (Scatter Plots)* використовують для відображення зв'язку між двома показниками для кожної статі, наприклад, залежності рівня доходів від рівня освіти. Для їх створення слід обрати дані й перейти на вкладку «Вставити» → «Розкид».

7. *Теплові карти (Heat Maps)* використовуються для демонстрації розподілу інтенсивності даних, наприклад, рівня зарплати чи оцінок задоволеності для різних статей у різних підрозділах. Для створення слід виділити дані й застосувати умовне форматування з «Кольорові шкали». Excel автоматично застосує кольорову шкалу до значень.

8. *Зведені діаграми та зведені таблиці* дозволяють агрегувати та фільтрувати дані за різними категоріями (стать, вік, посада) та швидко створювати відповідні діаграми. Для створення зведеної таблиці слід обрати потрібні поля для відображення та візуалізувати за допомогою зведеної діаграми.

Для ефективної візуалізації гендерних даних бажано дотримуватися таких рекомендацій:

А) завжди додавати підписи даних і легенди, щоб спростити сприйняття інформації.

Б) слід обирати кольори для відмінності між гендерними групами (наприклад, синій для чоловіків, червоний для жінок).

В) слід уникати надмірної деталізації, особливо у випадку секторних діаграм або складених діаграм, щоб не перевантажувати графік інформацією.

Отже, візуалізація гендерних даних надає можливість краще зрозуміти тенденції та забезпечити наочність для подальшого

аналізу. Розглянемо практичні способи обчислення простих гендерних індексів на прикладі таблиці з даними (рис. 6.2).

	A	B	C	D	E
1	Дані	Кіровоградська область	Україна	Європа + Північна Америка	Увесь світ
2	number_sph (f)	202 700	9 855 331	198 563 366	1 344 090 821
3	number_sph (m)	226 100	10 914 912	240 024 669	2 071 653 076
4	population (f)	447 189	20 876 809	389 932 647	2 750 154 246
5	population (m)	366 546	17 273 031	357 943 440	2 750 152 789
6	GPI	1,220	1,209	1,089	1,000
7	AGG	-80 643	-3 603 778	-31 989 207	-1 457
8	GPA	0,099	0,094	0,043	0,000
9	ratio (f)	0,453	0,472	0,509	0,489
10	ratio (m)	0,617	0,632	0,671	0,753
11	GPI (sph)	0,735	0,747	0,759	0,649
12	AGG (sph)	16,356	15,983	16,134	26,455
13	GPA (sph)	-0,153	-0,145	-0,137	-0,213

Рисунок 6.2. Вихідні дані та результати обчислення простих гендерних індексів.

Проведемо аналіз даних та визначимо індекси гендерного паритету, гендерного розриву та гендерної асиметрії за правилами знаходження таких індексів в демографічній та соціальній сфері.

Перші 4 рядки таблиці заповнені вихідними даними: рядки 2 та 3 – кількість осіб одного з регіонів відповідно жіночої та чоловічої статі, які є носіями певної функції (властивості, особливості) в соціальній сфері, рядки 4 та 5 – загальна кількість населення одного з цих регіонів. У стовпчику кожного з регіонів – Кіровоградська область, України, Європа плюс Північна Америка, увесь світ – пораховані:

- індекс гендерного паритету населення регіону (6 рядок, формула 6.1);
- індекс гендерного розриву (7 рядок, формула 6.2);
- індекс гендерної асиметрії (8 рядок, формула 6.3);
- співвідношення між кількістю осіб одного з регіонів жіночої (чоловічої) статі, які є носіями певної функції

(властивості, особливості) в соціальній сфері, до загальної кількості осіб жіночої (чоловічої) статі цього регіону (рядок 9 (10), формули 6.4);

- індекс питомого гендерного паритету (рядок 11, формула 6.5);
- індекс питомого гендерного розриву (рядок 12, формула 6.6);
- індекс питомої гендерної асиметрії (рядок 13, формула 6.7).

Прокоментуємо особливості застосування Excel для проведення відповідного аналізу простих гендерних у Кіровоградській області. Для знаходження індексу гендерного паритету населення регіону слід ввести формулу $=B4/B5$, індексу гендерного розриву $=B5-B4$, індексу гендерної асиметрії $=(B4-B5)/(B4+B5)$, співвідношення між кількістю осіб жіночої статі, які є носіями певної функції (властивості, особливості) в соціальній сфері, до загальної кількості осіб жіночої статі цього регіону $=B2/B4$, співвідношення між кількістю осіб чоловічої статі, які є носіями певної функції (властивості, особливості) в соціальній сфері, до загальної кількості осіб чоловічої статі цього регіону $=B3/B5$, індексу питомого гендерного паритету $=B9/B10$, індексу питомого гендерного розриву $=(B10-B9)*100$, індексу питомої гендерної асиметрії $=(B9-B10)/(B9+B10)$. Для інших регіонів значення відповідних індексів обчислюються аналогічно.

Перейдемо до роботи зі складними гендерними індексами. Розглянемо приклад з обчислення індексу гендерної нерівності (ГІ) та вивчення динаміки його зміни протягом 2010–2017 рр. у Кіровоградській області. Індекс ГІ відображає не тільки неблагополучне становище жінок в розширенні прав і можливостей, але і в сфері репродуктивного здоров'я. Також він відображає нерівність на рівні здобуття середньої освіти. Структура вимірів і індикаторів індексу ГІ є такою:

Вимір 1 – Здоров'я. Індикатори: а) рівень материнської смертності; б) підліткова народжуваність.

Вимір 2 – Розширення прав і можливостей. Індикатори: а) жіноче та чоловіче населення з принаймні середньою освітою; б) частки депутатських місць жінок і чоловіків.

Вимір 3 – Ринок праці. Індикатор: а) показники участі жінок і чоловіків у робочій силі.

Скориставшись даними Державної служби статистики України, демографічного щорічника «Населення України», інформацією сайту Кіровоградської обласної ради, Головного управління статистики у Кіровоградській області заповнимо конкретними даними таблицю 6.5.

Таблиця 6.5.

Складові показники індексу GII для Кіровоградської області

Рік	Здоров'я		Розширення прав та можливостей				Ринок праці	
	ABR	MMR	RC female	RC male	SE female	SE male	LFPR female	LFPR male
2010	38,6	85,5	16,16	83,84	94,2	92,1	58,1	67,4
2011	39,6	28,4	16,16	83,84	94,2	92,1	57,8	69,6
2012	42,2	18,1	16,16	83,84	94,2	92,1	57,8	71
2013	39,5	0	16,16	83,84	94,2	92,1	56	69
2014	40,9	37,8	16,16	83,84	94,2	92,1	55,2	67,5
2015	41,5	10,6	14,75	85,25	94,2	92,1	54,4	68,5
2016	34,1	33,9	14,75	85,25	94,2	92,1	54	67,5
2017	29,5	0	14,75	85,25	94,2	92,1	54,5	67,7

Проведемо обчислення GII Кіровоградської області в кілька етапів.

Етап 1. Облік нульових та екстремальних значень. Коефіцієнт материнської смертності (MMR) симетрично будемо обмежувати відрізком від 10 (мінімальне значення) до 1 000 (максимальне значення) для того, щоб підкреслити нормативність підходу. При реалізації такого підходу мається на увазі, що дві країни, які мають коефіцієнт материнської смертності вище 1 000, не відрізняються в нездатності створювати умови та організовувати турботу про здоров'я матерів; з іншого боку – приймається, що країни (області) з

1-10 випадками материнської смертності на 100 000 новонароджених функціонують на тому самому рівні в контексті попередження материнської смертності.

Представництво жінок у парламентах (обласних радах) країн (областей) (PR), зафіксоване на рівні 0%, кодуватимемо як 0,1%; це пов'язано з тим, що: а) середнє геометричне неспроможна мати нульове значення; б) жінки в таких країнах (областях) все ж таки мають деякий політичний вплив.

Етап 2. Агрегування вимірювань у межах кожної гендерної групи з використанням середніх геометричних величин. Агрегування, що спочатку проводиться за вимірюваннями всередині кожної гендерної групи з використанням середніх геометричних величин, є першим кроком на шляху до надання ГІ чутливості до взаємозв'язку. Обчислюється за формулами: для жінок формула агрегації така:

$$G_F = \sqrt[3]{\left(\frac{10}{MMR} \cdot \frac{1}{ABR}\right)^{1/2} \cdot (SE_F \cdot RC_F)^{1/2} \cdot LFPR_F}, \quad (6.12)$$

а для чоловіків:

$$G_M = \sqrt[3]{1 \cdot (SE_M \cdot RC_M)^{1/2} \cdot LFPR_M}. \quad (6.13)$$

Етап 3. Агрегування за гендерними групами з використанням гармонійної середньої величини. Індекси для жінок та чоловіків агрегують за середнім гармонійним (HARM) з метою створення рівнорозподіленого гендерного індексу, обчислюється за формулою

$$HARM(G_F, G_M) = \left[\frac{(G_F)^{-1} + (G_M)^{-1}}{2} \right]^{-1}. \quad (6.14)$$

Застосування середнього гармонійного значення середніх геометричних величин усередині груп враховує нерівність між жінками та чоловіками, одночасно коригуючи показник з урахуванням взаємозв'язку між вимірами.

Етап 4. Розрахунок середнього геометричного середніх арифметичних величин кожного індикатора. Референтний стандарт для розрахунку нерівності отримано шляхом агрегування чоловічого

та жіночого індексів з використанням рівних ваг (тобто при однаковому відношенні до обох статей), з подальшим агрегуванням цих індексів за вимірюваннями, обчислюється за формулами:

$$\overline{health} = \frac{\left(\sqrt{\left(\frac{10}{MMR}\right) \cdot \left(\frac{1}{ABR}\right)} + 1 \right)}{2}, \quad (6.15)$$

$$\overline{emp} = \left(\sqrt{SE_F \cdot RC_F} + \sqrt{SE_M \cdot RC_M} \right) \cdot 1/2, \quad (6.16)$$

$$\overline{LabourMarker} = (LFPR_F + LFPR_M) \cdot 1/2, \quad (6.17)$$

$$G_{F,M} = \sqrt[3]{\overline{health} \cdot \overline{emp} \cdot \overline{LabourMarker}}, \quad (6.18)$$

Величину $G_{F,M}$ називають референтним стандартом до розрахунку нерівності.

Показник \overline{health} слід розглядати не як середнє між відповідними чоловічим та жіночим індексами; а як половину відстані від норм, встановлених для індикаторів репродуктивного здоров'я – меншої кількості материнських смертей та меншої кількості вагітностей серед підлітків.

Етап 5. Розрахунок індексу гендерної нерівності. Порівняння рівнорозподіленого гендерного індексу з референтним стандартом дозволяє нам отримати Індекс гендерної нерівності, величина якої коливається від 0 (гендерна нерівність за всіма вимірами відсутня) до 1 (повна гендерна нерівність за всіма вимірами). Обчислюється за формулою:

$$GII = 1 - \frac{HARM(G_F, G_M)}{G_{F,M}}. \quad (6.19)$$

Такі операції слід провести з усіма даними таблиці 6.5. В результаті отримаємо показники GII для кожного року з періоду 2010–2017 рр. Динаміка зміни показників GII протягом вказаного періоду зображена на рис. 6.3.

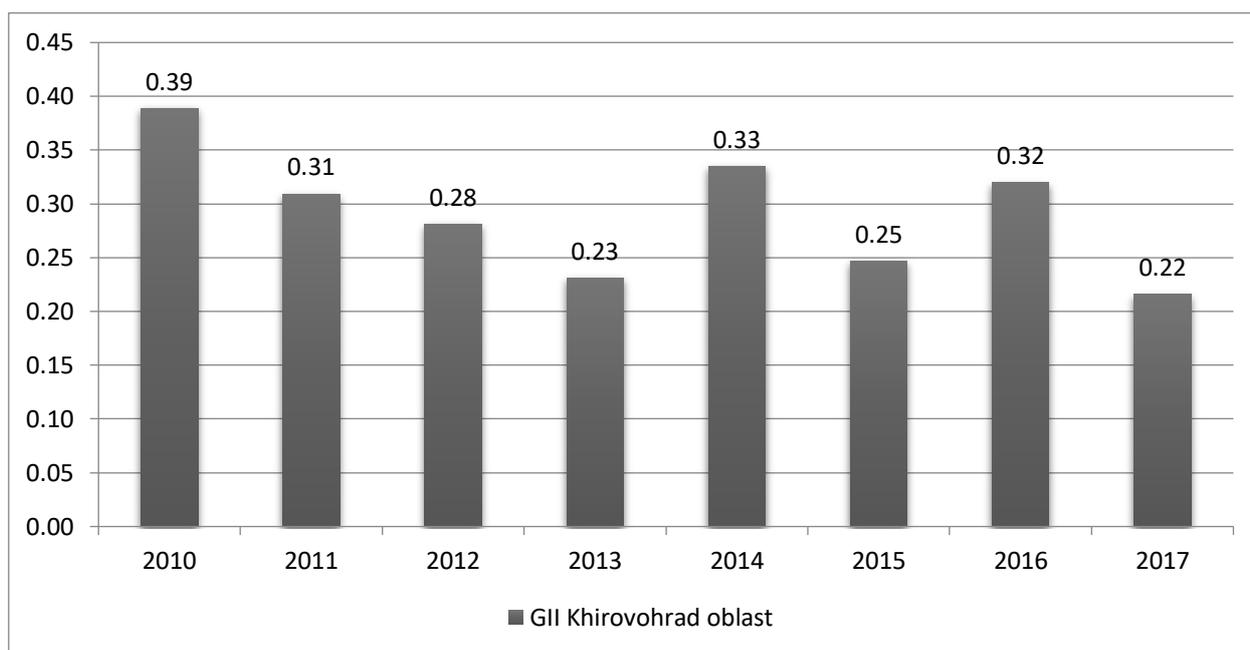


Рисунок 6.3. Динаміка зміни індексу гендерної рівності (обчислено за базовою методикою, 2010–2017 рр.)

Розглянемо практичне впровадження проведеної реалізації обрахунку GII в пакеті MSExcel.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1	Рік	Здоров'я		Розширення прав і можливостей				Ринок праці		GII	Gf	Gm	Harm	Health	Emp	Labour Marker	Gfm
2		AFR	MMR	PR female	PR male	SE female	SE male	LFPR female	LFPR male								
3	2010	38,6	85,5	0,1616	0,838	0,915	0,961	0,581	0,674	0,39	0,23	0,85	0,36	0,53	0,64	0,63	0,60
4	2011	39,6	28,4	0,1616	0,838	0,915	0,961	0,578	0,696	0,31	0,28	0,85	0,42	0,55	0,64	0,64	0,61
5	2012	42,2	18,1	0,1616	0,838	0,915	0,961	0,578	0,71	0,28	0,29	0,86	0,44	0,56	0,64	0,64	0,61
6	2013	39,5	0	0,1616	0,838	0,915	0,961	0,56	0,69	0,23	0,49	0,85	0,62	1,30	0,64	0,63	0,80
7	2014	40,9	37,8	0,1616	0,838	0,915	0,961	0,552	0,675	0,34	0,26	0,85	0,39	0,54	0,64	0,61	0,60
8	2015	41,5	10,6	0,1475	0,853	0,917	0,959	0,544	0,685	0,25	0,31	0,85	0,46	0,58	0,64	0,61	0,61
9	2016	34,1	33,9	0,1475	0,853	0,943	0,96	0,54	0,675	0,32	0,27	0,85	0,40	0,55	0,64	0,61	0,60
10	2017	29,5	0	0,1475	0,853	0,945	0,956	0,545	0,677	0,22	0,51	0,85	0,64	1,42	0,64	0,61	0,82

Рисунок 6.4. Робоча таблиця Excel для обрахунку динаміки GII Кіровоградської області (2010–2017 рр.)

Розглянемо особливості формування робочої таблиці на прикладі обрахунку GII (за базовою методикою) Кіровоградської області у 2017. При заповненні значення комірки K10 використовується формула (6.12), яка мовою пакета є такою: $=(\text{КОРЕНЬ}((10/0,1)*(1/V10))*\text{КОРЕНЬ}(D10*F10)*H10)^{(1/3)}$. При заповненні комірки L10 – формула (6.13) і запис $=(\text{КОРЕНЬ}(E10*G10)*I10)^{(1/3)}$. Для комірки M10 – формула (6.14) і запис $=(((1/K10)+(1/L10))/2)^{(-1)}$. Для комірки N10 – формула (6.15)

і запис $=(\text{КОРЕНЬ}((10/0,1)*(1/V10))+1)/2$. Для комірки O10 – формула (6.16) і запис $=(\text{КОРЕНЬ}(D10*F10)+\text{КОРЕНЬ}(E10*G10))/2$. Для комірки P10 – формула (6.17) і запис $=(H10+I10)/2$. Для комірки R10 – формула (6.18) і запис $=(N10*O10*P10)^{(1/3)}$. Нарешті, для заповнення комірки J10 використовується формула (6.19) і вноситься запис $=1-M10/Q10$.

Таким чином, використання Excel для аналізу гендерних індексів може бути ефективним, так як Excel є широко доступним і знайомим інструментом, що не вимагає від аналітиків значного навчання чи інвестицій у дорогі програми для виконання базових і навіть деяких складних аналітичних завдань. Інструменти Excel дозволяють швидко налаштувати процес збору, сортування, та фільтрації даних для різних категорій та показників, зокрема гендерних. Використання обчислювальних функцій забезпечує зручне обчислення гендерних індексів для різних гендерних груп. Функції організації порівнянь дозволяють співставляти дані, виконувати умовний підрахунок і порівнювати показники для різних груп на високому рівні точності та деталізації.

Інструменти візуалізації даних в Excel, такі як діаграми, гістограми та умовне форматування, дозволяють наочно представляти гендерні показники та тренди, що спрощує аналіз та прийняття рішень. Це робить результати більш зрозумілими для ширшої аудиторії та полегшує порівняння даних між гендерними групами. Умовне форматування підвищує зручність роботи з великими таблицями, дозволяючи швидко ідентифікувати відхилення або виділяти важливі дані.

Хоча Excel ефективний для базового та середнього рівня аналізу, він може мати певні обмеження при роботі з дуже великими обсягами даних або при проведенні складних статистичних розрахунків, що може вимагати спеціалізованих інструментів аналізу. Програма також не має вбудованих засобів для проведення більш глибоких статистичних тестів (наприклад, регресійного аналізу з великою кількістю змінних), що може

обмежити точність і глибину аналізу для деяких складних гендерних індексів.

Але в цілому Excel є ефективним інструментом для аналізу простих та складних гендерних індексів у межах середнього обсягу даних. Його функції обчислення, візуалізації та автоматизації дозволяють ефективно проводити аналітичну роботу і представляти її результати в зручному та доступному форматі. Проте для аналізу даних із вищим рівнем складності або для роботи з великими базами даних може знадобитися використання додаткових програм або спеціалізованих статистичних інструментів, таких як SPSS, R чи Python.

РОЗДІЛ 7.

РЕТРОПРОГНОЗУВАННЯ

7.1. Методологічні засади ретропрогнозування. Типи імітаційно-прогностичних моделей: альтернативна та контрфактична. Будування динамічних рядів за неповними даними

Ретропрогнозування – це методологічний підхід, що передбачає моделювання минулих подій або станів систем на основі наявних сучасних даних. Цей процес є важливою складовою аналітичних досліджень у багатьох галузях науки, таких як історія, економіка, кліматологія, геологія, біологія та соціологія. Його застосування дозволяє не лише відновлювати втрачену інформацію, але й перевіряти адекватність сучасних теоретичних моделей через порівняння їхніх результатів із задокументованими історичними фактами.

Метод ретропрогнозування є особливо актуальним у дослідженнях складних систем, де історичні дані часто є фрагментарними або неповними. Використовуючи сучасні математичні моделі, алгоритми машинного навчання та інші аналітичні інструменти, ретропрогнозування сприяє виявленню прихованих закономірностей і трендів, які можуть бути корисними для формування більш точних прогнозів щодо майбутнього розвитку системи.

Розділ присвячено аналізу принципів, методів і прикладів застосування ретропрогнозування, а також розгляду його

теоретичних і практичних обмежень. Основна увага приділяється ідентифікації чинників, які впливають на точність відтворення минулих подій, та оцінці ефективності методів у різних дослідницьких контекстах.

Ретропрогнозування та прогнозування є двома протилежними, але взаємопов'язаними методами аналізу та моделювання, що спрямовані на вивчення динаміки систем у часі. Основна відмінність між ними полягає в їхньому часовому векторі.

1. Ретропрогнозування та прогнозування мають різні *напрямки аналізу*. Прогнозування орієнтоване на майбутнє, воно використовує наявні дані про минуле та теперішнє для передбачення можливих сценаріїв розвитку системи в майбутньому. Ретропрогнозування спрямоване на минуле – цей метод використовує сучасні дані, моделі або інструменти для відновлення або реконструкції подій, що відбулися в минулому.

2. Ретропрогнозування та прогнозування різняться *метою*. Прогнозування має на меті надати інформацію для прийняття рішень або підготовки до майбутніх подій. Ретропрогнозування фокусується на перевірці теорій, моделей або уточненні історичних фактів шляхом порівняння отриманих результатів із реальними подіями минулого.

3. Ретропрогнозування та прогнозування різняться *наявністю фактів для перевірки*. При прогнозуванні результати важко перевірити до настання майбутнього. Їхня коректність може бути підтверджена або спростована лише з часом. При ретропрогнозуванні завжди є можливість перевіряти точність моделей, оскільки результати можна зіставити з відомими історичними даними.

4. Ретропрогнозування та прогнозування відрізняються *областями застосування*. Прогнозування широко використовується в економіці, метеорології, енергетиці, соціології, технологічних прогнозах тощо. Ретропрогнозування актуальне в історичних дослідженнях, археології, кліматології (моделювання минулого

клімату), геології (реконструкція тектонічних процесів), еволюційній біології (відтворення розвитку видів) тощо.

5. Ретропрогнозування та прогнозування відрізняються *інструментами та підходами реалізації*. Інструменти для ретропрогнозування і прогнозування часто є схожими (наприклад, математичні моделі, алгоритми машинного навчання). Однак для ретропрогнозування більшу роль відіграють методи верифікації даних із джерелами, що відображають реальні історичні події.

Отже, основна відмінність між ретропрогнозуванням та прогнозуванням полягає у фокусі на часі: прогнозування «заглядає вперед», тоді як ретропрогнозування «озирається назад». Однак обидва методи взаємодоповнюють одне одного, сприяючи глибшому розумінню складних систем у динаміці.

Ретропрогнозування як науковий підхід має глибокі історичні корені, що сягають перших спроб людства реконструювати минулі події та явища на основі доступних даних і спостережень. Його еволюція тісно пов'язана з розвитком наукових дисциплін, математичних методів і технологій обробки даних. Умовно історію появи та розвитку ретропрогнозування як наукового методу можна розділити на такі етапи.

1. *Ранні форми ретропрогнозування*. Перші спроби ретропрогнозування можна знайти в роботах античних мислителів, таких як Геродот чи Фулідд, які намагалися реконструювати історичні події на основі збережених письмових джерел і усних свідчень. У середньовіччі та ранньому новому часі ретропрогнозування отримало поштовх у галузі астрономії. Наприклад, Кеплер та Ньютон використовували математичні моделі руху небесних тіл для реконструкції їхніх позицій у минулому.

2. *Розвиток ретропрогнозування у природничих науках (XVIII–XIX ст.)*. У XVIII–XIX століттях ретропрогнозування почало активно застосовуватися в геології та палеонтології. Джеймс Гаттон (шотландський натураліст-геолог, основоположник плутонізму та уніформізму, його часто називають «батьком сучасної геології») та

Чарлз Лаелл (шотландський геолог, вулканолог, палеонтолог та гляціолог, основоположник сучасної геології; баронет, член Лондонського королівського товариства з розвитку знань про природу) розвинули ідею «актуалізму», припускаючи, що процеси, які формували Землю в минулому, аналогічні сучасним. Це дозволило реконструювати історію Землі та її тектонічних змін. У палеонтології вивчення скам'янілостей дало змогу відновлювати минулі екосистеми та хід еволюції видів.

3. *Становлення ретропрогнозування у соціальних і гуманітарних науках (XIX–XX ст.).* У соціології, економіці та історії ретропрогнозування отримало розвиток завдяки зростанню інтересу до кількісного аналізу та моделювання суспільних явищ. В економіці: ретроспективний аналіз економічних циклів допомагав виявляти закономірності для прогнозування криз і розвитку економіки. В історії застосування археологічних методів і технологій (наприклад, датування радіовуглецем) дозволило точніше реконструювати історичні події.

4. *Сучасний етап розвитку ретропрогнозування (XX–XXI ст.).* Сучасне ретропрогнозування сформувалося на основі досягнень у математичному моделюванні, статистиці, інформатиці та машинному навчанні. У другій половині XX століття розробка комп'ютерів дала змогу моделювати складні системи, включаючи ретроспективне моделювання клімату, екосистем і історичних процесів. Алгоритми штучного інтелекту, такі як нейронні мережі, дозволяють аналізувати величезні обсяги даних і будувати точніші моделі для реконструкції минулого. Інтеграція знань із різних галузей (фізики, біології, соціології, історії) зробила ретропрогнозування універсальним методом, який застосовується для вирішення широкого кола завдань – від реконструкції кліматичних змін до моделювання розвитку культур.

Ретропрогнозування пройшло тривалий шлях від простих спроб реконструювати минуле до складних математичних і комп'ютерних моделей. Його сучасний розвиток сприяє глибшому

розумінню як минулого, так і майбутнього, роблячи цей метод важливим інструментом для науки та суспільства.

На сьогоднішній день ретропрогнозування як науковий метод знаходить застосування у різноманітних галузях, що потребують реконструкції минулих подій, процесів чи станів систем. Воно дозволяє аналізувати складні системи на основі сучасних даних, з метою відтворення їхньої еволюції. Перелічимо основні галузі, де цей метод активно використовується.

1. *Застосування в природничих науках.* В геології ретропрогнозування застосовується для реконструкції геологічної історії Землі (зокрема руху тектонічних плит, утворення континентів і океанів), для вивчення вулканічної активності та визначення її впливу на клімат у минулому, з метою аналізу стратиграфічних даних для відновлення осадових умов у давні епохи.

В кліматології ретропрогнозування застосовується з метою ретроспективного моделювання клімату для дослідження змін у температурі, атмосферному складі, льодовиковому покриві та рівні океанів, а також для відновлення кліматичних умов на основі даних із кернів льоду, коралів, деревних кілець і осадових порід.

В палеонтології ретропрогнозування служить методом реконструкції екосистем минулих епох на основі скам'янілостей та аналізу еволюції видів, вивчення біогеографічних змін та визначення причин масових вимирань.

2. *В історії та археології* ретропрогнозування застосовується з метою відновлення історичних подій за архівними, археологічними та письмовими джерелами, для реконструкції стародавніх поселень, інфраструктур, торгових шляхів і культурних зв'язків, для аналізу демографічних змін у давніх цивілізаціях, а також для датування історичних об'єктів і артефактів за допомогою методів, таких як радіовуглецеве датування.

3. *В астрономії та космології* ретропрогнозування застосовується для визначення положення небесних тіл у минулому з метою вивчення змін орбітальних параметрів, для реконструкції історії

Сонячної системи (зокрема формування планет і поясу астероїдів), для аналізу даних із наднових зірок для встановлення їхньої ролі в еволюції галактик.

4. *Застосування в біології та екології.* В еволюційній біології ретропрогнозування застосовується для проведення ретроспективного аналізу геномів для вивчення еволюції видів і популяцій та для визначення спільних предків видів на основі філогенетичного аналізу.

В екології цей метод застосовується для реконструкції змін у біорізноманітті та екосистемах під впливом кліматичних чи антропогенних факторів, а також для моделювання динаміки зміни обсягу популяцій у минулому.

5. *Застосування у соціальних науках.* В економіці ретропрогнозування застосовується для проведення аналізу економічних криз, бізнес-циклів і довгострокових трендів економічного розвитку, а також для здійснення ретроспективного моделювання фінансових ринків для виявлення причин ринкових аномалій.

У соціології ретропрогнозування є важливим при організації дослідження демографічних змін, міграційних потоків і соціальної стратифікації у минулому, а також для вивчення впливу соціальних потрясінь (революцій, війн) на суспільства.

6. *В медицині та епідеміології* ретропрогнозування використовується при аналізі поширення минулих епідемій та оцінка їхніх наслідків для здоров'я популяцій, для реконструкції генетичних мутацій вірусів і бактерій для розуміння їхньої еволюції, для відтворення історичних медичних практик з метою вивчення їхньої ефективності.

7. *В техніці та технології* ретропрогнозування застосовується для організації ретроспективного аналізу розвитку інфраструктур (транспорт, енергетика) з метою оцінки їхньої ефективності в різні періоди, а також для відновлення технологічних процесів минулого для збереження традиційних методів виробництва.

8. Нарешті, *в праві та криміналістиці* метод ретропрогнозування є незамінним при реконструкції злочинів на основі аналізу доказів (наприклад, ДНК, слідів на місці події), а також для організації аналізу юридичних документів з метою відтворення правових систем у минулому.

Галузі застосування ретропрогнозування продовжують розширюватися завдяки зростанню обчислювальної потужності сучасної комп'ютерної техніки та реалізації доступу до великих даних. У майбутньому метод може стати ще потужнішим інструментом для вирішення міждисциплінарних задач, які вимагають реконструкції минулого з точністю та деталізацією.

Ретропрогнозування як науковий метод базується на низці основних принципів, які забезпечують його ефективність і надійність. Ці принципи формують основу для моделювання, аналізу й інтерпретації даних з метою реконструкції минулих подій чи станів систем.

1. *Принцип причинно-наслідкових зв'язків* у ретропрогнозуванні означає, що будь-яка подія чи явище в минулому є результатом певних причин і взаємодій. Ретропрогнозування передбачає виявлення цих причин та їхнього впливу на розвиток системи. Припущення, що подібні причини у схожих умовах призводять до однакових наслідків, допомагає побудувати моделі, здатні відтворювати минуле.

2. *Принцип актуалізму*, який особливо поширений у геології та палеонтології, стверджує, що процеси, які діяли в минулому, є такими ж, як і в сучасності. Наприклад, знання про сучасні вулканічні чи ерозійні процеси використовується для розуміння геологічної історії Землі.

3. *Принцип збереження енергії та матерії* про застосуванні ретропрогнозування означає, що фізичні закони, зокрема закон збереження енергії, матерії чи імпульсу, є ключовими для ретропрогнозування в природничих науках. Цей принцип дозволяє встановити межі можливих сценаріїв реконструкції минулого, наприклад, у геодинаміці чи астрономії.

4. *Принцип історичної контекстуалізації* означає, що ретропрогнозування враховує контекст часу та простору, в якому відбувалися події. Наприклад, у соціальних науках реконструкція суспільних процесів базується на врахуванні політичних, економічних і культурних умов епохи.

5. *Принцип ймовірності та статистичного аналізу* при застосуванні ретропрогнозування означає, що з огляду на неповноту та неточність історичних даних, ретропрогнозування використовує ймовірнісні підходи для оцінки можливих сценаріїв. Наприклад, методи байєсівського аналізу чи Монте-Карло дозволяють визначити ступінь невизначеності результатів.

6. *Принцип оберненості моделей* означає, що ретропрогнозування передбачає використання сучасних моделей і даних для «зворотного» моделювання – повернення до станів, які передували спостережуваним. Це особливо важливо у фізичних і математичних задачах, наприклад, у моделюванні динаміки клімату або еволюції космічних систем.

7. *Принцип перевірки гіпотез* означає, що побудовані моделі ретропрогнозування повинні бути перевірені шляхом порівняння з незалежними джерелами даних (архівні матеріали, скам'янілості, керни льоду тощо). Це забезпечує верифікацію моделей та підвищує їхню точність.

8. *Принцип міждисциплінарності* значить, що ретропрогнозування вимагає залучення знань і методів із різних галузей науки (математики, фізики, хімії, історії, біології тощо). Поєднання підходів дозволяє отримати більш повну картину минулого, наприклад, через інтеграцію археологічних і геохімічних даних.

9. *Принцип сталості природних законів* передбачає, що закони природи (наприклад, гравітація, термодинаміка) були однаковими в минулому, як і сьогодні. Це дає змогу застосовувати сучасні теоретичні й експериментальні знання для відновлення подій минулого.

10. *Принцип адаптивності моделей* означає, що моделі ретропрогнозування мають бути гнучкими та здатними до уточнення на основі нових даних. Постійне оновлення баз даних та методологій сприяє підвищенню точності результатів.

Перелічені принципи забезпечують наукову обґрунтованість і надійність ретропрогнозування. Їх дотримання дозволяє не лише реконструювати минулі події, а й використовувати отримані знання для кращого розуміння сучасних процесів і прогнозування майбутнього.

Ретропрогнозування базується на кількох основних підходах, які забезпечують ефективність реконструкції минулих подій та явищ. Ці підходи залежать від специфіки об'єкта дослідження, доступності даних та інструментів аналізу. Ключові підходи, на яких ґрунтується ретропрогнозування, є такими.

1. *Реалізація математичного моделювання* – цей підхід передбачає створення математичних моделей, що описують поведінку систем у часі. При реалізації підходу використовуються детерміновані моделі, які використовують фізичні або математичні закони для обчислення стану системи в минулому (наприклад, рівняння Нав'є-Стокса для моделювання кліматичних процесів) та стохастичні моделі, що враховують невизначеність і випадкові фактори, які могли впливати на систему (наприклад, для моделювання популяційних динамік або фінансових ринків).

2. *Реалізація реконструктивного підходу*, який базується на відтворенні минулих подій шляхом аналізу наявних даних та об'єктів. При цьому розрізняють археологічну реконструкцію – аналіз матеріальних залишків для відновлення історичних подій, побуту або інфраструктур та палеоекологічну реконструкцію – використання природних маркерів, таких як пилок, осадові породи чи керни льоду, для відтворення минулих екосистем чи клімату.

3. *Реалізація комп'ютерного моделювання* шляхом застосування алгоритмів і симуляцій для ретроспективного аналізу складних систем у вигляді симуляції на основі великих даних – сучасні суперкомп'ютери дозволяють обробляти величезні обсяги даних

для моделювання історичних процесів, наприклад, міграцій чи економічних змін. Другим варіантом є геопросторове моделювання – використання геоінформаційних систем (ГІС) для аналізу змін у ландшафті чи географічному розташуванні об'єктів у часі.

4. Реалізація *статистичного підходу* дозволяє виявити тренди, закономірності та кореляції у ретроспективних даних. При цьому використовуються регресійний аналіз для встановлення залежностей між різними факторами (наприклад, між кліматичними змінами та динамікою популяцій) та байєсівські моделі, які дозволяють оцінювати ймовірність минулих подій на основі сучасних даних.

5. Реалізація *історико-аналітичного підходу*, який полягає у вивченні та зіставленні історичних джерел для реконструкції подій. При такому підході використовується текстуальний аналіз стародавніх текстів, літописів і документів для отримання даних про минуле та порівняльний аналіз для зіставлення даних із різних джерел для перевірки точності реконструкції.

6. Реалізація *експериментального підходу* для відтворення минулих умов у контрольованих експериментальних умовах для перевірки гіпотез. Для цього використовуються лабораторні (наприклад, моделювання хімічних реакцій, що могли відбуватися в первинній атмосфері Землі) та польові експерименти (реконструкція стародавніх технологій або методів господарства для перевірки їхньої ефективності).

7. Реалізація *генетичного підходу* при ретропрогнозуванні у галузі біології та медицини, що базується на аналізі генетичної інформації. Застосовується при цьому молекулярна філогенетика – вивчення геномів для визначення еволюційних зв'язків і спільних предків та аналіз ДНК – реконструкція історії популяцій, шляхів міграції або поширення захворювань.

8. Реалізація *геохімічного підходу* для аналізу хімічного складу природних матеріалів, щоб реконструювати минулі процеси. При цьому використовується ізотопний аналіз – визначення віку

об'єктів чи процесів (наприклад, через радіовуглецеве датування) та вивчення хімічних маркерів – аналіз слідів забруднень чи змін у геологічних породах.

9. Реалізація *моделювання обернених задач*, що застосовується в фізиці, математиці та інженерії для розв'язання задач, де відомий результат, але невідомі початкові умови. Наприклад, у сейсмології обернені задачі дозволяють визначати джерело землетрусу на основі даних про сейсмічні хвилі.

10. Реалізація *інтердисциплінарного підходу*, що являє собою поєднання методів і підходів із різних дисциплін, яке дозволяє отримати більш повну картину минулого (наприклад, інтеграція даних з археології, генетики та кліматології для реконструкції історії міграції людства).

Отже, ретропрогнозування використовує широкий спектр підходів, які можуть бути адаптовані до специфіки конкретної галузі. Інтеграція цих підходів дозволяє з високою точністю реконструювати минулі події, забезпечуючи як фундаментальне розуміння, так і практичне застосування у сучасних дослідженнях.

Ретропрогнозування історичних подій і процесів спирається на широкий спектр джерел історичних даних. Їх можна класифікувати залежно від типу інформації, яку вони надають, і методу використання в дослідженнях.

При цьому використовуються *письмові джерела* – архівні документи, офіційні записи (закони, укази, судові протоколи, економічні звіти), особисті документи (листи, щоденники, мемуари), адміністративні документи (переписи населення, звіти про господарську діяльність), історичні хроніки та літописи (які містять деталі про політичні, релігійні, військові та природні події), літературні твори (поєми, драми, оповідання можуть містити інформацію про суспільний устрій, культуру та повсякденне життя), газети та журнали (відображають події, громадську думку та соціальні тенденції певного часу).

Крім письмових використовуються *археологічні знахідки* (будівлі, кераміка, знаряддя праці, прикраси, монети; вони дають

інформацію про рівень технологічного розвитку, торгівлю, економіку, культуру), артефакти з місць поховань (знаряддя праці, предмети ритуального значення, останки тварин і людей, вони відновлюють уявлення про соціальну стратифікацію, релігійні уявлення та побут), твори мистецтва – малюнки, скульптури, настінні розписи (містять візуальну інформацію про культуру, ідеологію та повсякденне життя).

Використовуються також *природні джерела* – палеокліматичні дані (керни льоду дають інформацію про температуру, склад атмосфери, концентрацію газів у минулому), дані дендрохронологічного аналізу (аналіз річних кілець дерев для встановлення кліматичних умов), вивчення осадових порід (хімічний склад і шари осадів дають дані про природні умови), дані аналізу біологічних залишків (пилки, насіння, кістки, ДНК давніх організмів, які використовуються для реконструкції екосистем, раціону та середовища існування).

Цінними є *іконографічні джерела* – карти, схеми, гравюри, ілюстрації (відображають географічне положення, містобудування, культурні особливості), фотографії та кіноматеріали для більш сучасних періодів.

Важливими також є *усні джерела* – легенди, перекази, фольклор (особливо цінні для відновлення історії народів, які не мали письмової традиції), усні історії сучасників певних подій.

Часто використовуються у якості джерел *географічні дані* – геоінформаційні системи (дані про розташування стародавніх поселень, ландшафти, маршрути міграцій; вони цінні для проведення аналізу змін ландшафтів через карти та топографічні дані), аерофотозйомка та супутникові знімки (використовуються для виявлення археологічних об'єктів та змін ландшафту).

Використовуються також *генетичні дані* – давня ДНК (аналіз геномів людини, рослин, тварин для вивчення еволюційних процесів, міграцій, змін у популяціях), ізотопний аналіз зубів і кісток (для виявлення раціону, походження та умов життя).

Важливими у якості джерел є *метеорологічні та астрономічні дані* – стародавні астрономічні записи (спостереження за сонячними затемненнями, кометами, зірками для встановлення точних дат подій), природні явища (описи землетрусів, вивержень вулканів чи повеней у літописах).

Важливим є також використання *технічних джерел* – стародавніх інструментів, машин, механізмів, які дають уявлення про технологічний розвиток і виробничі процеси.

Нарешті, цінними є *міждисциплінарні джерела* – інтеграція різних джерел (поєднання письмових, археологічних і природних даних для комплексного аналізу; наприклад, співставлення кліматичних даних із письмовими свідченнями про врожайність чи голод).

Отже, джерела історичних даних для ретропрогнозування охоплюють різноманітні аспекти людського життя, природного середовища та культурного розвитку. Їх комплексне використання дозволяє відтворити минулі події з високою точністю та деталізацією.

Обробка та аналіз історичних даних для ретропрогнозування є багатоступеневим процесом, який включає збирання, оцінку, інтерпретацію та моделювання даних. Цей процес вимагає використання міждисциплінарних підходів, що забезпечують комплексне відтворення минулих подій.

На першому етапі проводиться *збір даних*, який включає ідентифікацію джерел (вибір релевантних письмових, археологічних, природних чи інших даних залежно від досліджуваного питання; наприклад, для аналізу змін клімату можуть бути використані керни льоду, а для реконструкції міграцій – археологічні артефакти й генетичні дані), оцифрування та каталогізацію (оцифрування архівних документів, фотографій, карт і артефактів; організація даних у структуровані бази даних для подальшого аналізу) та залучення додаткових джерел (використання супутникових знімків, геоінформаційних систем, лабораторних зразків тощо).

На другому етапі проводиться *попередня обробка даних*, що передбачає очистку даних (видалення дублюючої інформації, помилок і «шуму»; наприклад, виправлення розбіжностей у датах або назвах у письмових джерелах), стандартизацію форматів (перетворення даних у формат, придатний для комп'ютерного аналізу – таблиці, бази даних, цифрові карти), вирівнювання часових шкал між джерелами (наприклад, об'єднання даних за однаковими часовими періодами), нормалізацію (приведення даних до спільної одиниці вимірювання або масштабу; наприклад, нормалізація економічних показників за інфляцією чи приведення координат до однієї системи).

Третій етап передбачає проведення *інтерпретації даних*. Тут проводиться контекстуалізація (аналіз історичного, соціального, політичного та економічного контексту для правильного трактування даних; наприклад, врахування воєн чи природних катаклізмів при інтерпретації змін демографічних даних), хронологічна реконструкція (упорядкування подій і фактів за часовою шкалою для виявлення динаміки змін), використання методів датування (радіовуглецеве, ізотопне, дендрохронологія), порівняння джерел (зіставлення даних із різних джерел для перевірки їхньої достовірності та заповнення прогалін).

На четвертому етапі проводиться *аналіз даних*, де передбачається організація статистичного аналізу (використання регресійного, кореляційного та кластерного аналізу для виявлення закономірностей і залежностей у даних, байєсівського аналізу для оцінки ймовірності різних сценаріїв минулого), моделювання обернених задач (використання математичних моделей для відновлення початкових умов на основі результатів; наприклад, для реконструкції клімату чи міграцій), проведення геоінформаційного аналізу (застосування геоінформаційних систем для аналізу географічних змін і просторових взаємозв'язків; наприклад, вивчення змін ландшафту чи міграцій шляхом моделювання шляхів пересування), застосування симуляції та прогнозування (використання комп'ютерних симуляцій для

моделювання сценаріїв розвитку подій; наприклад, аналіз можливих причин соціального колапсу на основі історичних даних).

На п'ятому етапі проводиться *верифікація та уточнення моделей*, що проводиться шляхом порівняння з новими даними (верифікація отриманих моделей через співставлення з нововиявленими джерелами чи даними; наприклад, зіставлення результатів моделювання з археологічними відкриттями), тестування гіпотез (перевірка коректності моделей за допомогою незалежних методів – експериментальні відтворення, аналіз альтернативних сценаріїв), уточнення моделей (оновлення моделей із урахуванням нових даних чи виправлення помилок у припущеннях).

Шостий етап роботи з історичними даними для ретропрогнозування – *інтерпретація результатів та візуалізація*. Він включає візуалізацію даних (побудову графіків, карт, діаграм і 3D-реконструкцій для наочного представлення результатів; наприклад, створення карт історичних міграцій або кліматичних змін), формулювання висновків (узагальнення результатів для реконструкції минулих подій, аналізу закономірностей та прогнозування майбутнього), публікацію результатів (викладення результатів дослідження у вигляді звітів, статей, книг чи інтерактивних презентацій).

Нарешті, на останньому етапі проводиться *міждисциплінарна інтеграція*, яка передбачає поєднання історичних даних із даними з біології, геології, економіки чи соціології, створення комплексної картини минулого, яка враховує всі аспекти досліджуваного питання.

Таким чином, обробка та аналіз історичних даних у ретропрогнозуванні є складним і багатоетапним процесом, який забезпечує високу точність реконструкції минулого. Використання сучасних інструментів і міждисциплінарних підходів дозволяє інтегрувати різноманітні джерела даних для створення цілісної картини історичних подій і процесів.

Для ретропрогнозування використовуються різноманітні моделі, які допомагають реконструювати події минулого, враховуючи специфіку об'єкта дослідження та доступні дані. Основні моделі можна класифікувати за їх характером та галуззю застосування.

1. *Фізичні моделі*, які базуються на законах фізики і застосовуються для реконструкції природних процесів. Вони включають кліматичні моделі, що використовуються для відновлення минулих кліматичних умов (наприклад, моделі загальної циркуляції атмосфери відновлюють клімат на основі даних із кернів льоду чи осадів), тектонічні моделі, які відновлюють зміни земної поверхні (наприклад, рухи тектонічних плит або вулканічну активність), гідрологічні моделі, що використовуються для аналізу змін водних систем (ріки, озера) в минулому (наприклад, рівнів води в річках або льодовикових періодів).

2. *Статистичні моделі*, що базуються на аналізі історичних даних для виявлення закономірностей і прогнозування минулих процесів. Вони включають регресійні моделі, які використовуються для встановлення залежностей між змінними (наприклад, між демографічними даними та економічними показниками), байєсівські моделі, що оцінюють ймовірність певних подій на основі сучасних даних і ймовірнісних припущень про минуле, кластерні моделі, що групують об'єкти або явища для виявлення схожих закономірностей у минулому (наприклад, у демографічних чи археологічних даних).

3. *Моделі обернених задач*, які реконструюють початкові умови системи, використовуючи відомі результати. Вони включають сейсмологічні моделі, призначені для визначення параметрів минулих землетрусів (використовують сейсмічні хвилі, зафіксовані сучасними пристроями), геофізичні моделі, які реконструюють розподіл магнітних або гравітаційних аномалій у минулому.

4. *Біологічні та генетичні моделі*, що використовуються для аналізу минулої еволюції та поширення живих організмів. Вони включають філогенетичні моделі, що відновлюють еволюційні

зв'язки між видами на основі аналізу ДНК або викопних решток), популяційні моделі, що реконструюють міграції людей або тварин, зміни чисельності популяцій у минулому, екологічні моделі, що відновлюють зміни екосистем на основі даних про біорізноманіття чи сліди людської діяльності.

5. *Економічні моделі*, які використовуються для аналізу розвитку економічних систем у минулому. Вони включають моделі економічного зростання, які допомагають оцінити розвиток економіки в історичних періодах (наприклад, вплив технологічних інновацій), торговельні моделі, які відтворюють стародавні торговельні шляхи та обсяги торгівлі між регіонами, демографо-економічні моделі, що аналізують взаємозв'язок між демографічними змінами та економічною активністю.

6. *Геоінформаційні моделі*, які включають моделі просторового аналізу для реконструкції змін ландшафту, розташування поселень, міграційних маршрутів; моделі розподілу населення, які аналізують, як змінювалося розташування та густина населення на територіях у минулому.

7. *Комп'ютерні симуляційні моделі*, що створюють комп'ютерні симуляції минулих подій. Вони включають агент-орієнтовані моделі, які відтворюють взаємодію індивідів або груп у соціальних, економічних чи екологічних системах; системно-динамічні моделі, що аналізують динаміку складних систем (наприклад, соціальних чи економічних структур).

8. *Історико-аналітичні моделі*, що працюють на основі текстових даних (використовують текстовий аналіз історичних документів для реконструкції подій; наприклад, методи обробки природної мови (NLP) дозволяють автоматично аналізувати письмові джерела) та реалізують синтез історичних джерел шляхом зіставлення інформації з літописів, архівів і фольклору для створення комплексних моделей подій.

9. *Соціокультурні моделі*, які включають моделі соціальної динаміки, які вивчають вплив соціальних факторів (міграцій, воєн, революцій) на розвиток суспільства, та культурно-історичні моделі,

які відтворюють зміни в культурних традиціях, мовах і релігіях на основі археологічних та письмових джерел.

10. Нарешті, *інтегровані міждисциплінарні моделі*, до складу яких входять комплексні моделі (об'єднують підходи з кількох галузей – екології, археології, історії, генетики – для створення цілісної картини минулого), прогностично-реконструктивні моделі (використовують сучасні знання та тенденції для точнішої реконструкції минулого).

Окремо відзначимо особливості реалізації *моделей машинного навчання* (МН), які активно використовуються в ретропрогнозуванні для аналізу великих обсягів даних, виявлення прихованих закономірностей і створення точних реконструкцій минулого. Завдяки адаптивності і здатності працювати з неповними або неоднорідними даними, МН стають важливим інструментом у цій галузі.

При підготовці даних для МН використовуються структуровані й неструктуровані дані: історичні записи, археологічні матеріали, кліматичні серії, супутникові зображення, генетичні дані тощо. Попередня обробка включає оцифрування, очищення, нормалізацію, злиття даних з різних джерел. Позначені дані (наприклад, кліматичні дані з відомими параметрами) використовуються для тренування моделей. Непозначені дані аналізуються з використанням методів кластеризації чи факторного аналізу.

Розрізняють такі варіанти застосування методів машинного навчання в ретропрогнозуванні:

1. *Моделі для регресійного аналізу*: а) лінійна регресія та її розширення, оцінка залежностей між змінними, наприклад, між демографічними показниками і природними факторами; б) рішення дерев і випадкових лісів, прогнозування на основі послідовних розгалужень за різними факторами, наприклад, визначення причин змін клімату.

2. *Моделі кластеризації*: а) К-середніх: групування археологічних артефактів або об'єктів за подібними ознаками; б) DBSCAN

(Density-Based Spatial Clustering): виділення просторових груп даних, наприклад, поселень або місць поховань.

3. *Глибоке навчання* (Deep Learning): а) рекурентні нейронні мережі (RNN): використовуються для аналізу часових рядів (температур, економічних змін, демографії); б) згорткові нейронні мережі (CNN): аналіз зображень, наприклад, стародавніх карт, супутникових знімків або фотографій артефактів.

4. *Ненаглянуті моделі*: використовуються для аналізу історичних даних без попереднього маркування. Наприклад, автоенкодері допомагають знайти приховані структури в археологічних комплексах або культурних артефактах.

5. *Методи Байєсівського виведення*: допомагають оцінити ймовірність певних сценаріїв минулого на основі сучасних даних і припущень. Наприклад, використовуються для реконструкції стародавніх екосистем.

Серед галузей застосування МН у ретропрогнозуванні називають кліматичне ретропрогнозування (МН аналізує дані із шарів льоду, осадових порід і деревних кілець, щоб відновити кліматичні умови тисячоліть тому; симуляція погодних умов для виявлення довготривалих змін), геопросторовий аналіз (МН моделі виявляють закономірності у просторовому розташуванні історичних об'єктів; наприклад, визначення місць можливих поселень або торгівельних маршрутів), аналіз текстових даних (Natural Language Processing (NLP) використовується для аналізу стародавніх текстів, автоматичного перекладу, виявлення ключових подій або понять у літописах), археологічну реконструкцію (аналіз даних розкопок для прогнозування можливих місць знаходження артефактів; створення 3D-моделей історичних об'єктів або структур), еволюцію та генетику (моделі допомагають відтворити еволюційні шляхи видів або міграцію людських популяцій, використовуючи генетичні дані), соціально-економічне ретропрогнозування (аналіз соціальних і економічних змін на основі демографічних даних, стародавніх економічних записів;

визначення причин колапсів цивілізацій або злету економічної активності).

Отже, ретропрогнозування спирається на багатий арсенал моделей, кожна з яких відповідає конкретним аспектам історичних досліджень. Поєднання цих моделей забезпечує глибше розуміння минулого й дозволяє реконструювати складні процеси з високою точністю.

Розглянемо, як математико-статистичні методи використовуються при організації ретропрогнозування. Математичні та статистичні методи є основою аналізу минулих даних і побудови ретропрогнозів. Вони дозволяють формалізувати залежності між змінними, оцінювати ймовірність подій і будувати моделі, що реконструюють історичні явища. Нижче розглянемо основні підходи та методи.

1. *Статистичний аналіз* історичних даних використовує: а) дескриптивну статистику, яка описує ключові характеристики даних: середні значення, медіани, моду, дисперсію, стандартне відхилення; дескриптивна статистика використовується для попередньої оцінки історичних даних, таких як температури, рівень води чи демографічні показники; б) кореляційний аналіз, який допомагає виявити зв'язки між змінними. Наприклад, як зміни клімату впливали на рівень сільськогосподарської продуктивності. При цьому використовуються такі методи: коефіцієнт кореляції Пірсона, Спірмена, Кендалла; в) розподіли ймовірностей, при якому аналізуються закономірності розподілу даних (нормальний, експоненціальний, Пуассонів). Наприклад, для оцінки частоти виникнення природних катаклізмів у минулому.

2. *Регресійний аналіз* дозволяє визначити залежність між змінними та прогнозувати значення однієї змінної на основі іншої. Використовується: а) лінійна регресія, яка визначає лінійний зв'язок між змінними і використовується для аналізу простих залежностей, наприклад, між кількістю населення та обсягом вирощеного зерна; б) поліноміальна регресія, яка використовується для аналізу нелінійних залежностей, наприклад, зміни температури протягом

століть; в) багатofакторна регресія, яка аналізує взаємозв'язки між кількома факторами, наприклад, вплив клімату, економічних змін і соціальних факторів на розвиток суспільств.

3. *Аналіз часових рядів* дозволяє досліджувати дані, впорядковані в часі. При цьому актуальними є такі дії: а) виділення довгострокових трендів і циклічних змін у даних, наприклад, кліматичних або демографічних (застосовується декомпозиція часових рядів, рухоме середнє, експоненційне згладжування); б) моделі ARIMA (AutoRegressive Integrated Moving Average) – моделі для аналізу й прогнозування часових рядів з урахуванням автокореляції (використовуються для реконструкції змін клімату, річкових потоків або економічних показників); в) спектральний аналіз, який виявляє приховані цикли в часових рядах, наприклад, цикли сонячної активності або кліматичні цикли.

4. *Теорія ймовірностей і моделювання* передбачає використання: а) байєсівського підходу, який використовується для оцінки ймовірності подій з урахуванням нових даних (наприклад, для оцінки ймовірності соціальних конфліктів у давніх суспільствах); є важливим для ситуацій, коли дані неповні або нечіткі; б) Монте-Карло моделювання – використовує випадкові вибірки для моделювання складних процесів і оцінки ймовірності певних сценаріїв (наприклад, для аналізу поширення епідемії або зміни популяції).

5. *Аналіз просторових даних* допомагає відновити географічний розподіл явищ. При цьому застосовуються: а) кригінг – метод інтерполяції, який дозволяє прогнозувати значення в невідомих точках на основі даних сусідніх точок; використовується для реконструкції ландшафтів, ґрунтових характеристик або кліматичних умов; б) аналіз щільності – допомагає визначити концентрацію явищ, наприклад, поселень або місць археологічних знахідок.

6. Для аналізу великих обсягів історичних даних застосовуються *методи кластеризації, факторного аналізу та видобування даних*. При цьому використовується: а) кластеризація –

групування об'єктів за схожими характеристиками, наприклад, поселень за типами забудови (використовуються методи К-середніх, DBSCAN); б) аналіз головних компонент (РСА) – зменшує розмірність даних для виявлення основних факторів впливу (наприклад, виділення ключових кліматичних або соціальних факторів).

7. *Інтердисциплінарні підходи*, які включають: а) комбіновані моделі, які поєднують статистичні методи з іншими, наприклад, моделюванням фізичних процесів (гідрологічних, кліматичних); б) верифікацію моделей – перевірку моделей на сучасних даних або подібних історичних періодах для підвищення точності прогнозів.

Отже, математичні та статистичні методи забезпечують системний підхід до аналізу минулих даних, дозволяючи будувати обґрунтовані ретропрогнози. Вони дають змогу оцінювати залежності, виявляти закономірності та моделювати складні явища, що є ключовими для глибокого розуміння минулого.

У ретропрогнозуванні ключову роль відіграють комп'ютерні технології та програмне забезпечення, забезпечуючи обробку великих обсягів даних, автоматизацію аналізу та моделювання складних історичних процесів. Вони розширюють можливості дослідників, дозволяючи виконувати високоточні обчислення, інтегрувати дані з різних джерел і створювати візуалізації для розуміння минулого.

1. Комп'ютерні технології забезпечують ефективну *роботу з даними* на всіх етапах ретропрогнозування:

- Оцифрування історичних джерел: сканування та розпізнавання текстів, карт, археологічних даних.
- Створення баз даних: структуроване збереження інформації для подальшого аналізу (наприклад, історичні бази даних кліматичних показників (Paleoclimatology Database)).

- Програмне забезпечення для обробки текстів і зображень: Adobe Photoshop для обробки археологічних зображень, ABBYY FineReader для розпізнавання тексту.

2. Комп'ютерні технології уможливають проведення автоматизованого аналізу даних:

- Програмне забезпечення для статистичного аналізу – пакет SPSS, мови програмування R, Python (бібліотеки: pandas, NumPy, SciPy) забезпечують аналіз статистичних даних, реалізують побудову моделей регресії, кластеризацію; пакет MATLAB реалізує обробку складних математичних моделей і часових рядів.
- Обробка великих даних реалізується з використанням технологій Big Data (таких як Apache Hadoop і Spark) для обробки масивів історичних даних, отриманих із супутників, археологічних баз, літописів.
- Геоінформаційні системи (ГІС), які реалізуються програмним забезпеченням (наприклад, ArcGIS, QGIS) використовуються для аналізу просторових даних (географічних особливостей, міграційних шляхів), побудови карт історичних поселень і реконструкції змін ландшафту.
- Інструменти обробки зображень (наприклад, ImageJ, Photoshop, AutoCAD) реалізують аналіз зображень археологічних знахідок, 3D-реконструкцію артефактів.

3. Комп'ютерні технології розв'язують задачі, пов'язані з моделюванням історичних процесів:

- Комп'ютерне моделювання забезпечує моделювання соціально-економічних систем і демографічних змін (наприклад, AnyLogic, Vensim); продукує симуляцію кліматичних умов у минулому (наприклад, Climate models (GCMs)).
- Машинне навчання реалізує використання нейронних мереж, рекурентних моделей (RNN, LSTM) для аналізу

часових рядів (наприклад, інструменти: TensorFlow, PyTorch, Scikit-learn).

- Агентно-орієнтоване моделювання розв'язує задачу симуляції поведінки індивідуальних агентів у історичних контекстах (моделювання міграцій, соціальної взаємодії; наприклад, NetLogo, Repast).

4. Комп'ютерні технології дозволяють *візуалізувати складні дані* для кращого розуміння (наприклад, Data visualization tools, Tableau, Power BI, Python (Matplotlib, Seaborn), зображати графіки змін температури, населення, економічних показників, реалізовувати 3D-моделювання (наприклад, Blender, SketchUp для реконструкції будівель, ландшафтів), анімувати історичні зміни – створювати динамічні карти і відео, що показують еволюцію подій.

5. Комп'ютерні технології *розв'язують задачі*:

- Автоматизації системи збору та аналізу даних у варіанті використання веб-скрейпінгу для збору даних з архівів і онлайн-баз, або у вигляді автоматизації аналізу текстів історичних документів через Natural Language Processing (NLP).
- Розробки систем підтримки прийняття рішень (DSS), які використовують для вибору найкращих підходів до моделювання, інтерпретації даних і оцінки сценаріїв ретропрогнозування.

6. Комп'ютерні технології відкривають можливості *впровадження штучного інтелекту (ШІ)*:

- Використання ШІ для розпізнавання закономірностей у великих даних і автоматичного створення моделей ретропрогнозу.
- Аналіз текстів для визначення ключових подій і понять у літописах, хроніках.
- Переклад стародавніх мов і автоматичний аналіз змісту.

7. Реалізуються *хмарні технології та онлайн-платформи*:

- Хмарні сервіси (Google Cloud, AWS) забезпечують доступ до великих обчислювальних потужностей і обміну даними між дослідниками.
- Організація онлайн-платформ для співпраці (наприклад, Open Science Framework (OSF)), дозволяють публікувати та обговорювати результати.
- Хмарні сховища для інтеграції даних з різних джерел, наприклад, архівів, бібліотек і супутникових баз.

Але слід пам'ятати, що використання у ретропрогнозуванні комп'ютерних технологій має й певні виклики, які можуть бути пов'язані з неповнотою даних (навіть передові алгоритми потребують якісних вихідних даних), складністю моделей (розуміння виходів складних моделей може бути проблемою для нематематичних спеціалістів), етичними питаннями (обмеження використання технологій у певних регіонах або з чутливими даними).

Отже, комп'ютерні технології та програмне забезпечення суттєво розширюють межі ретропрогнозування, забезпечуючи точність, швидкість і зручність аналізу. Вони дають змогу дослідникам відтворювати минуле на основі інтеграції даних, математичного моделювання та сучасних технологій візуалізації, сприяючи більш глибокому розумінню історичних процесів.

Аналіз часових рядів є одним із ключових етапів у ретропрогнозуванні, адже він дозволяє виявити закономірності, тренди, циклічні коливання та інші важливі аспекти, пов'язані з подіями минулого. Спеціалізовані програмні пакети значно полегшують цю роботу, забезпечуючи зручний інтерфейс, багатий набір функцій і алгоритмів для аналізу. Розглянемо основні особливості та приклади їх використання.

1. *Основні функції спеціалізованих пакетів* для аналізу часових рядів є такими:

- Усунення шуму та вибірок із часового ряду.
- Обробка відсутніх даних (інтерполяція, заповнення середніми значеннями).

- Перетворення даних (нормалізація, логарифмування).
- Визначення середніх значень, дисперсії, автокореляції.
- Виявлення сезонності, циклічності, трендів.
- Моделювання часових рядів шляхом побудови моделей ARIMA, SARIMA, ETS (Exponential Smoothing), Prophet.
- Прогнозування на основі минулих даних.
- Побудова графіків часових рядів, трендів, спектрального аналізу.
- Оцінка точності моделей (метрики: RMSE, MAE, MAPE).
- Перевірка на наявність автокореляції залишків (тест Дікі-Фуллера, Ljung-Box).

2. Можна назвати такі *приклади спеціалізованих пакетів* і їх застосування:

- Мова програмування Python, у якій пропонується широкий спектр бібліотек для роботи з часовими рядами: а) Pandas – робота з часовими даними, фільтрація, агрегування, ресемплінг (наприклад: аналіз економічних показників у ретроспективі); б) Statsmodels – побудова моделей ARIMA, SARIMA, VAR – використовується для аналізу кліматичних даних, таких як температурні аномалії; в) Prophet (Meta) – модель із вбудованим урахуванням трендів, сезонності та впливу свят яка використовується для аналізу довгострокових даних, наприклад, населення або врожайності; г) SciPy та NumPy – фільтрація даних, спектральний аналіз.
- Мова програмування R, яка є однією з провідних платформ для аналізу часових рядів: а) forecast – побудова моделей ETS, ARIMA, прогнозування сезонних змін, аналіз економічних і кліматичних даних; б) tseries – інструменти для тестів на стаціонарність (тест Дікі-Фуллера), автокореляції, для перевірки моделей на адекватність; в) zoo та xts – робота з нерівномірними часовими рядами (наприклад, дані археологічних розкопок); г) ggplot2 – візуалізація даних часових рядів.

- Пакет MATLAB, який широко використовується для аналізу наукових часових рядів, таких як кліматичні чи гідрологічні: а) Econometrics Toolbox – моделювання ARIMA, VAR, GARCH для економічних даних; б) Signal Processing Toolbox – спектральний аналіз, фільтрація.
 - Пакети Tableau та Power BI, що використовуються для інтерактивної візуалізації часових рядів.
 - Геоінформаційні системи (ГІС): а) ArcGIS, QGIS –: аналіз часових змін у просторі, наприклад, міграційні процеси або зміни ландшафту.
3. Застосування спеціалізованих пакетів у ретропрогнозуванні
- Реконструкція кліматичних змін: а) використання моделей ARIMA у Python або R для аналізу даних про температури, рівень опадів; б) спектральний аналіз для виявлення кліматичних циклів (наприклад, циклів Ель-Ніньо).
 - Економічні ретропрогнози: а) аналіз історичних даних про ВВП, торговельному обігу; б) Моделі ETS або Prophet для визначення впливу економічних криз.
 - Археологічні та демографічні дослідження: а) аналіз часових рядів із баз археологічних знахідок для реконструкції історичних періодів; б) виявлення циклічності міграційних процесів.
 - Ретроспективний аналіз природних катастроф: а) моделювання даних про землетруси, повені або посухи; б) прогнозування періодичності подібних явищ на основі ретропрогнозів.

Таким чином, спеціалізовані пакети для аналізу часових рядів значно спрощують ретропрогнозування, дозволяючи точно аналізувати минулі події, моделювати їх динаміку та прогнозувати можливі сценарії. Їх використання у поєднанні з великими обсягами даних дає змогу дослідникам відтворювати історичні процеси з високою точністю та інтегрувати результати у сучасний науковий контекст.

Ретропрогнозування як науковий метод стикається з низкою обмежень і викликів, які ускладнюють його використання та впливають на точність і достовірність результатів. Ці проблеми пов'язані з якістю даних, методологічними аспектами, технічними складнощами та етичними питаннями.

По-перше, обмеження можуть бути пов'язані з *якістю даних*, а саме з фрагментарністю історичних джерел (значна частина даних можуть бути втрачені або неповні через знищення документів, артефактів, природні катаклізми), обмеженістю часового горизонту (більшість доступних даних охоплюють лише останні кілька століть, що ускладнює аналіз давніших періодів). Обмеження також можуть бути пов'язані з низькою точністю даних – історичні записи часто містять помилки через людський фактор, застарілі методи вимірювань або суб'єктивність описів; археологічні дані можуть мати значні похибки датування (наприклад, метод радіовуглецевого аналізу має обмежену точність). Нарешті, обмеження можуть бути спричинені відсутністю стандартизації – дані з різних регіонів і епох можуть бути несумісними через відмінності в методах збору й запису інформації.

По-друге, ретропрогнозування може стикатися з *методологічними викликами*: а) вибір моделі (вибір невідповідної моделі для аналізу може призвести до викривлення результатів; багато моделей припускають лінійність процесів, тоді як історичні явища часто нелінійні та хаотичні); б) стаціонарність даних (більшість моделей аналізу часових рядів вимагають стаціонарності, тобто незмінності середнього значення та дисперсії у часі; проте багато історичних даних є нестаціонарними через вплив глобальних змін (наприклад, війни, природні катаклізми)); в) інтерпретація результатів (отримані моделі потребують складної інтерпретації, щоб уникнути помилкових висновків; потрібно враховувати культурний, соціальний і політичний контекст, який може бути недостатньо зрозумілим через обмеженість даних).

По-третє, ретропрогнозування може стикатися з *технічними викликами*: а) робота з великими даними (аналіз великих масивів

даних потребує високих обчислювальних потужностей і складних алгоритмів; інтеграція різних типів даних (текстових, числових, географічних) може викликати технічні труднощі); б) обмеженість програмного забезпечення (не всі доступні інструменти мають потрібну функціональність для складного аналізу історичних даних; адаптація сучасних методів (наприклад, машинного навчання) до специфіки ретропрогнозування може вимагати додаткової розробки); в) чутливість моделей до параметрів (моделі ретропрогнозування часто дуже чутливі до вибору початкових умов і параметрів, що може призводити до значних похибок у прогнозах).

По-четверте, ретропрогнозування може стикатися з *етичними викликами*: а) використання історичних даних (використання архівних матеріалів може бути обмежене через авторські права, політичні чи культурні причини; використання даних, що стосуються чутливих тем (війни, геноцид), потребує обережного підходу); б) викривлення історичних фактів (неправильна інтерпретація або моделювання може призвести до створення неточних реконструкцій історичних подій; потенційна маніпуляція результатами для виправдання сучасних ідеологічних чи політичних цілей).

Ретропрогнозування як науковий метод може стикатися з *іншими обмеженнями*, а саме: а) часові та фінансові ресурси (ретропрогнозування є трудомістким і дорогим процесом, який потребує великих інвестицій у дослідження, обладнання та розробку програмного забезпечення); б) висока міждисциплінарність (успішне ретропрогнозування вимагає знань у різних галузях: історії, статистиці, географії, комп'ютерних науках, що створює труднощі у формуванні міждисциплінарних команд).

Є різні *шляхи подолання цих викликів та обмежень*. У якості найбільш дієвих назовемо такі:

- Інтеграція сучасних технологій – використання штучного інтелекту, хмарних технологій та інструментів Big Data.

- Співпраця між дослідниками – створення міжнародних проєктів, які об'єднують ресурси й експертизу.
- Розробка нових методологій – адаптація сучасних математичних моделей до специфіки історичних досліджень.
- Підвищення якості даних – оцифрування архівів, створення централізованих баз даних, удосконалення методів збору інформації.

Отже, хоча ретропрогнозування має значний потенціал для розуміння історичних процесів, воно стикається з численними викликами. Їх подолання можливе завдяки поєднанню сучасних технологій, міждисциплінарного підходу та міжнародної співпраці. Це дозволить зробити ретропрогнозування ще ефективнішим інструментом для наукових досліджень.

У ретропрогнозуванні існує досить серйозний ризик, який полягає у схильності дослідника шукати, інтерпретувати або використовувати історичні дані таким чином, щоб вони підтверджували вже існуючі гіпотези чи очікування. Це явище називається ризик підтверджувального зміщення (confirmation bias). Це може суттєво спотворювати результати, знижувати об'єктивність та обмежувати можливість відкриття нових закономірностей.

Що таке підтверджувальне зміщення? Це когнітивна помилка, за якої: а) дані, що відповідають очікуванням, отримують більше уваги й ваги; б) суперечливі дані можуть ігноруватися, інтерпретуватися як несуттєві або «аномалії»; в) аналіз та моделювання підпорядковуються підтвердженню бажаного результату.

Як підтверджувальне зміщення проявляється у ретропрогнозуванні? Це може траплятися кількома способами: а) при виборі даних – відбір тільки тих джерел чи часових рядів, які підтверджують гіпотезу; б) при інтерпретації даних – надання переваги теоріям, які відповідають припущенням дослідника, навіть за наявності суперечливих доказів; в) при побудові моделей –

налаштування параметрів моделей так, щоб вони відповідали очікуваним результатам; г) ігнорування невизначеностей – недооцінка похибок або впливу обмеженості даних.

Наведемо такі приклади підтверджувального зміщення в ретропрогнозуванні.

- При проведенні ретроспективного аналізу економічних криз використання тільки тих економічних показників, які підтверджують теорію про зовнішні причини кризи, ігноруючи можливі внутрішні фактори.
- При проведенні аналізу кліматичних змін відбір тільки даних, які підтримують концепцію глобального потепління, ігноруючи дані про локальні кліматичні аномалії.
- При проведенні археологічних досліджень зосередження уваги на знахідках, які підтверджують ідею про існування конкретної цивілізації, нехтуючи суперечливими матеріалами.

Для уникнення ризику підтверджувального зміщення можуть використовуватися такі засоби та прийоми:

- Використання прозорої методології – визначення чітких критеріїв вибору даних до початку аналізу; повідомлення про всі етапи дослідження, включаючи відхилені дані чи моделі.
- Залучення альтернативних підходів – перевірка гіпотез за допомогою кількох незалежних моделей; використання різних джерел даних і різних типів аналізу для підтвердження результатів.
- Усвідомлення когнітивних упереджень – постійна критична оцінка власних припущень і методів; залучення незалежних експертів для рецензування дослідження.
- Аналіз протилежних даних – активний пошук даних, які можуть суперечити основній гіпотезі; ретельний аналіз «аномалій» для з'ясування їхньої природи.

- Використання автоматизованих інструментів – використання машинного навчання та нейтральних алгоритмів, які мінімізують вплив суб'єктивних рішень дослідника; наприклад, при побудові моделей часових рядів автоматизовані підходи (наприклад, ARIMA, Prophet) можуть обрати оптимальні параметри без втручання дослідника.
- Застосування міждисциплінарного підходу – робота в команді фахівців із різних галузей, що дозволяє уникнути обмеженості однієї точки зору.

Суттєву допомогу в уникненні ризику підтверджувального заміщення може надати процедура перевірки результатів: а) тестування на різних наборах даних – застосування моделі до даних із різних періодів, регіонів чи джерел; перевірка стабільності результатів при зміні вхідних параметрів; б) проведення «сліпого тестування» – аналіз даних без попереднього знання їхнього походження або контексту; в) оцінка похибок і невизначеностей – розрахунок довірчих інтервалів для результатів моделей; використання метрик для порівняння точності різних моделей.

Отже, підтверджувальне зміщення є суттєвим викликом у ретропрогнозуванні, адже воно може призводити до спотворення історичної реконструкції. Уникнення цього ризику можливе завдяки прозорій методології, критичному підходу до аналізу, використанню сучасних автоматизованих технологій і співпраці міждисциплінарних команд. Це забезпечує підвищення достовірності результатів і сприяє отриманню об'єктивних наукових висновків.

Ретропрогнозування як науковий метод знаходить широке застосування в різних галузях завдяки своїй здатності аналізувати минулі події та відтворювати їхні причинно-наслідкові зв'язки. Це дає можливість не лише зрозуміти сутність історичних процесів, а й спрогнозувати можливі варіанти розвитку ситуацій у схожих умовах. Успіх методу базується на обробці великих обсягів даних,

використанні сучасних математичних моделей та потужних комп'ютерних технологій.

Приклади використання ретропрогнозування охоплюють такі ключові сфери, як економіка, кліматологія, демографія, археологія, медицина та багато інших. У кожній з цих галузей метод дозволяє реконструювати хід подій, які відбувалися за відсутності точних даних або за умов обмеженої інформації. Завдяки цьому можна розробляти більш обґрунтовані теорії, проводити оцінку ефективності минулих рішень і покращувати методи прийняття рішень у майбутньому.

Розглянемо конкретні приклади застосування ретропрогнозування у різних галузях, зокрема в економіці, де аналізуються причини та наслідки криз, довгострокове економічне зростання й ефективність політик, та інших сферах, де ретропрогнозування слугує незамінним інструментом для відновлення картини минулого.

Ретропрогнозування в економіці активно використовується для аналізу минулих подій, моделювання економічних процесів і створення сценаріїв розвитку, що допомагають зрозуміти причини змін та наслідки рішень у макро- і мікроекономіці. Наведемо приклади його практичного застосування.

1. Аналіз економічних криз.

При вивченні особливостей Великої депресії (1929–1939) ретропрогнозування використовується для вивчення причин обвалу фондового ринку, динаміки безробіття та впливу державних інтервенцій (наприклад, політики «Новий курс» у США). При цьому на основі реконструйованих часових рядів аналізуються ефекти фіскальної та монетарної політики, що дозволяє оцінити ефективність антикризових заходів.

При вивченні фінансової кризи 2008 року дослідники використовують ретропрогнозування, щоб моделювати, як різні сценарії втручання центральних банків (наприклад, зниження ставок, «кількісне пом'якшення») могли змінити хід подій. Аналіз

попередніх криз (наприклад, Азіатська криза 1997 року) слугує базою для ретропрогнозів щодо 2008 року.

2. Реконструкція економічного зростання.

При вивченні динаміки зміни ВВП у довгостроковій перспективі ретропрогнозування застосовується для аналізу економічного зростання країн за відсутності прямої статистики (наприклад, у Середньовіччі або ранньомодерний період). Наприклад, для реконструкції ВВП Британії у XVIII столітті використовуються дані про сільськогосподарське виробництво, ціни на зерно, податкові звіти тощо.

З'ясування особливостей економічного розвитку колоній, а саме дослідження економічної історії колоній (наприклад, Північної Америки, Індії) допомагає оцінити вплив колоніальної політики на торгівлю, фінанси та суспільний добробут.

3. Ретропрогнозування інфляційних процесів.

Моделі ретропрогнозування застосовуються для аналізу гіперінфляції у країнах, таких як Німеччина у 1920-х роках чи Зімбабве на початку 2000-х років. Використовуючи дані про обсяги грошової маси, зростання цін та рівень зарплат, дослідники можуть оцінити вплив різних факторів на інфляцію та визначити потенційні точки перетину критичних порогів.

4. Вивчення ефективності економічної політики.

При дослідженні політики протекціонізму ретропрогнозування дозволяє оцінити наслідки запровадження митних тарифів та квот у минулих періодах (наприклад, політика США у 1930-х роках). За допомогою моделей аналізується вплив таких заходів на зайнятість, виробництво та зовнішню торгівлю.

При аналізі податкової політики може бути проведена реконструкція впливу змін ставок оподаткування на доходи державного бюджету й економічне зростання. Наприклад, оцінка наслідків податкових реформ у Великій Британії XIX століття.

5. Аналіз ринкової динаміки.

При дослідженні фондових ринків моделювання минулих біржових крахів (наприклад, «чорний вівторок» 1929 року) дозволяє

зрозуміти динаміку спекулятивних бульбашок і роль психологічних факторів у поведінці інвесторів.

При проведенні аналізу сировинних ринків аналіз змін цін на нафту чи метали в періоди великих економічних потрясінь допомагає оцінити роль геополітичних факторів у формуванні ринкової кон'юнктури.

6. Оцінка соціально-економічного впливу катастроф.

Ретропрогнозування використовується про дослідженні пандемії. Наприклад, ретропрогнозування економічних наслідків пандемії «іспанського грипу» 1918 року допомагає оцінити вплив на зайнятість, доходи та промислове виробництво.

Ретропрогнозування використовується про дослідженні природних катаклізмів. Аналіз економічних втрат після землетрусів, повеней чи посух дозволяє оцінити вплив на сільське господарство, торгівлю й міграцію населення.

7. Моделювання альтернативних сценаріїв.

Що було б, якби... моделі – наприклад, оцінка альтернативного розвитку подій, якщо би країни вчасно приймали антикризові рішення (наприклад, впровадження «плану Маршалла» в інших регіонах).

Сценарне моделювання дозволяє оцінити довгострокові наслідки ключових економічних рішень.

Отже, ретропрогнозування в економіці є потужним інструментом для аналізу минулого й пошуку закономірностей, які можуть бути застосовані для прийняття сучасних рішень. Воно допомагає краще зрозуміти складні економічні явища, оцінити ефективність політик і підготуватися до потенційних викликів у майбутньому.

Ретропрогнозування в археології є ефективним інструментом для реконструкції минулого, моделювання розвитку цивілізацій і виявлення закономірностей у формуванні культурних ландшафтів. Завдяки використанню математичних моделей, географічних даних та комп'ютерного моделювання, дослідники можуть оцінювати вплив екологічних, соціальних та економічних факторів

на розвиток стародавніх суспільств. Наведемо приклади використання ретропрогнозування в археології.

1. Реконструкція стародавніх ландшафтів.

Ретропрогнозування використовується для відтворення географічних умов, які існували тисячі років тому. Наприклад, у регіоні Межиріччя моделюються зміни русел річок Тигр і Євфрат, які могли впливати на розташування міст-держав.

Використання даних про кліматичні умови (грунтуючись на аналізі пилку, осадових порід і льодовикових шарів) дозволяє зрозуміти, як посухи чи потепління впливали на сільське господарство та міграцію населення.

2. Дослідження демографічних змін.

Ретропрогнозування допомогло встановити залежність розміру населення від рівня розливів Нілу в період Давнього Єгипту. Моделі також враховували зміни в родючості ґрунтів і розвиток іригаційних систем.

Аналіз демографічних даних майя чи мешканців Анасазі в Північній Америці використовує ретропрогнозування, щоб пояснити, як перенаселення, деградація ґрунтів і зміни клімату призвели до зникнення цих культур.

3. Відтворення торговельних шляхів.

За допомогою ретропрогнозування реконструюються маршрути, якими проходили торговці в античному світі, враховуючи дані про рельєф, кліматичні умови та економічні зв'язки.

Аналіз морських шляхів на основі стародавніх карт і даних про знайдені корабельні уламки дозволяє відтворити мережу торгівлі між містами-державами Греції та Риму.

4. Реконструкція технологічного розвитку.

Ретропрогнозування використовується для визначення етапів розвитку технологій обробки заліза, бронзи чи золота. Моделювання умов видобутку та поширення металів допомагає зрозуміти, як ці технології впливали на економіку та суспільство.

Аналіз стародавніх будівель (наприклад, пірамід Єгипту чи римських акведуків) дозволяє відтворити методи їх будівництва, враховуючи обмеженість технічних ресурсів того часу.

5. Розуміння ритуальних практик.

Використання даних про географічне розташування поховань дозволяє моделювати соціальну ієрархію та ритуальні традиції. Наприклад, аналіз курганів скіфів у Причорномор'ї допомагає зрозуміти їхні поховальні звичаї.

Ретропрогнозування використовується для моделювання транспортних та будівельних технологій, які могли застосовуватися для спорудження Стоунхенджа чи подібних комплексів.

6. Дослідження землекористування.

У Мезоамериці моделі ретропрогнозування використовуються для відтворення традиційних систем землеробства майя, включаючи підняття терас і створення каналів для іригації.

Аналіз ґрунтових проб і даних про ерозію дозволяє визначити, як неправильне використання ресурсів могло сприяти занепаду давніх цивілізацій.

7. Вивчення катастроф та їх впливу на культуру.

Реконструкція наслідків виверження вулкана на Санторіні (Тіра) допомогла пояснити, як цей катаклізм вплинув на занепад мінойської цивілізації.

Аналіз змін Гангу та Інда в контексті культури Хараппи допомагає зрозуміти причини занепаду цієї цивілізації.

Отже, ретропрогнозування в археології допомагає не тільки відновити минуле, а й зробити його зрозумілим у контексті сучасних викликів. Використовуючи цей підхід, археологи отримують можливість перевіряти гіпотези, створювати моделі та відтворювати складні соціально-економічні та екологічні процеси стародавніх суспільств.

Ретропрогнозування в соціальних науках дозволяє вивчати минулі події та процеси, аналізувати соціальні зміни й моделювати їхні наслідки. Використовуючи історичні дані, математичні моделі й міждисциплінарний підхід, дослідники отримують глибше

розуміння того, як змінювалися суспільства та які фактори впливали на їхній розвиток. Наведемо приклади використання ретропрогнозування в соціальних науках.

1. Аналіз демографічних процесів.

Ретропрогнозування допомагає реконструювати чисельність населення в різні періоди історії, використовуючи дані про народжуваність, смертність та міграцію. Наприклад, моделювання демографічного переходу в Європі XVIII–XIX століть.

Аналіз наслідків «Чорної смерті» (1347–1351 рр.) для чисельності населення Європи з урахуванням соціально-економічних факторів, таких як міграція та зміни в землекористуванні.

2. Дослідження соціальної мобільності.

Використання ретропрогнозування для аналізу соціальної мобільності в історичних суспільствах. Наприклад, вивчення мобільності між кастами в Індії або класової структури в середньовічній Європі.

Аналіз того, як доступ до освіти впливав на соціальну мобільність у різних історичних епохах, наприклад, після запровадження обов'язкової освіти в Західній Європі.

3. Вивчення соціальних конфліктів.

Моделювання причин і наслідків революцій, таких як Французька революція чи Жовтнева революція в Росії, з урахуванням економічних, політичних та ідеологічних чинників.

Ретропрогнозування використовується для оцінки впливу великих воєн, наприклад, Першої світової війни, на соціальну структуру, зайнятість і права жінок.

4. Реконструкція економічної нерівності.

За допомогою ретропрогнозування оцінюється рівень економічної нерівності в історичних суспільствах. Наприклад, реконструкція рівня доходів у Римській імперії чи середньовічній Європі.

Аналіз наслідків запровадження різних податкових систем, таких як десятина в середньовіччі, для розподілу багатства.

5. Дослідження урбанізації.

Використання ретропрогнозування для вивчення урбанізаційних процесів, наприклад, під час індустріальної революції у Великій Британії чи формування мегаполісів у стародавніх цивілізаціях, таких як Теночтитлан (стародавнє місто, столиця ацтекської імперії, розташоване на островах посеред озера Тескоко в сучасній Мексиці; засноване близько 1325 року н.е. ацтеками (мешіками) і стало одним із найбільших і найрозвиненіших міст доколумбової Америки; на піку свого розвитку, перед прибуттям іспанських конкістадорів на чолі з Ернаном Кортесом у 1519 році, Теночтитлан мав населення, за оцінками, від 200 до 300 тисяч осіб, що робило його одним із найбільших міст у світі того часу).

Моделювання міграційних потоків до міст із сільської місцевості під впливом економічних і соціальних чинників.

6. Аналіз культурних змін.

Ретропрогнозування використовується для аналізу розповсюдження мов і культур, наприклад, латини в Римській імперії чи слов'янських мов у Європі.

Дослідження поширення релігій, таких як християнство чи іслам, і їхній вплив на соціальні та політичні структури.

7. Оцінка наслідків технологічних інновацій.

Ретропрогнозування використовується для моделювання впливу технологічних нововведень, таких як паровий двигун, на зайнятість, економіку та соціальні відносини.

Аналіз того, як впровадження нових агротехнологій впливало на соціальну організацію та демографію в різних регіонах.

8. Дослідження політичних систем.

Моделювання розвитку демократичних інститутів у країнах Європи після революцій XVIII–XIX століть.

Ретропрогнозування використовується для аналізу причин розпаду таких імперій, як Османська чи Радянська, із урахуванням соціальних та політичних змін.

Таким чином, ретропрогнозування в соціальних науках дозволяє глибше зрозуміти процеси, які формували сучасні суспільства. Завдяки міждисциплінарному підходу та використанню сучасних технологій, дослідники отримують можливість оцінити не лише наслідки минулих подій, а й побудувати прогнози для майбутнього на основі історичного досвіду.

Ретропрогнозування як науковий метод постійно розвивається, адаптуючись до вимог сучасних досліджень і нових технологій. Зі зростанням обсягів даних, доступних для аналізу, і появою більш потужних обчислювальних інструментів, методи ретропрогнозування набувають усе більшої точності й універсальності. Удосконалення цих методів дозволяє не лише якісніше реконструювати минулі події, але й розширювати їх застосування у різних наукових сферах, таких як економіка, соціологія, кліматологія та археологія.

Сучасні інновації охоплюють інтеграцію великих даних, використання штучного інтелекту, геопросторових технологій, вдосконалених математичних моделей та віртуальної реальності. Водночас розвиток інструментів для візуалізації та моделювання результатів відкриває нові можливості для комунікації результатів досліджень широкій аудиторії. Розглянемо ключові напрями інновацій у ретропрогнозуванні та перспективи вдосконалення його методології.

Інновації та вдосконалення методів ретропрогнозування спрямовані на підвищення точності моделювання, ефективності обробки даних та інтеграції нових технологій для аналізу складних історичних процесів. Наведемо основні напрями розвитку цієї галузі.

1. Використання великих даних (*Big Data*).

Інтеграція великих обсягів різнорідних даних – використання історичних записів, архівних матеріалів, супутникових даних, археологічних знахідок та екологічних показників для створення повнішої картини минулого.

Автоматизована обробка даних – інноваційні алгоритми дозволяють ефективніше обробляти текстові, числові й просторові дані, зменшуючи ризик помилок і збільшуючи точність моделей.

2. Розвиток машинного навчання та штучного інтелекту (AI).

Глибоке навчання (Deep Learning) – моделі штучного інтелекту здатні виявляти приховані закономірності в історичних даних, які складно помітити за допомогою традиційних методів.

Самонавчальні моделі – алгоритми, що адаптуються до нових даних, дозволяють поліпшувати точність ретропрогнозування з кожним наступним етапом аналізу.

Обробка неструктурованих даних – наприклад, аналіз історичних текстів або зображень для відновлення подій.

3. Використання геопросторових технологій (GIS).

Просторовий аналіз – геоінформаційні системи дозволяють моделювати ландшафтні, кліматичні та соціальні зміни, що впливали на історичні події.

Тривимірне моделювання – реконструкція міських структур, ландшафтів чи бойових дій на основі археологічних знахідок і документальних джерел.

4. Покращення математичних моделей.

Моделі з мультиагентним підходом – включення в моделі різних «агентів» (окремих людей, груп, країн) дозволяє детальніше аналізувати взаємодії в соціальних, економічних чи політичних системах.

Моделювання з урахуванням невизначеностей – створення моделей, які враховують нестачу даних чи їхню неоднозначність, з використанням ймовірнісного аналізу або байєсівських підходів.

5. Інтеграція з міждисциплінарними підходами.

Поєднання даних з археології, антропології, історії та екології – це дозволяє враховувати ширший контекст подій та їхній вплив на суспільство.

Кліматичні моделі – інтеграція даних про кліматичні зміни з соціальними та економічними показниками для реконструкції довгострокових процесів.

6. Віртуальна та доповнена реальність (VR/AR).

Візуалізація даних – використання VR/AR для відтворення історичних подій або умов життя дозволяє зрозуміти результати ретропрогнозування не лише дослідникам, а й широкій аудиторії.

Інтерактивні симуляції – можливість тестувати різні сценарії минулого та оцінювати альтернативні шляхи розвитку подій.

7. Хмарні обчислення та високопродуктивні обчислювальні системи.

Обробка великих обсягів даних у реальному часі – хмарні платформи дозволяють ефективно зберігати та обробляти дані, забезпечуючи доступ дослідникам з усього світу.

Високопродуктивні обчислювальні системи (HPC) – їх використання пришвидшує складні розрахунки, необхідні для ретропрогнозування на основі великих даних.

8. Нові підходи до валідації моделей.

Зворотне тестування – порівняння результатів ретропрогнозування з відомими історичними фактами для перевірки точності моделей.

Колаборативні дослідження – залучення експертів з різних дисциплін для критичного аналізу моделей і їхнього вдосконалення.

9. Етичні аспекти та прозорість.

Прозорість моделей – розробка зрозумілих алгоритмів і моделей, які можуть бути перевірені іншими дослідниками.

Етичний підхід – врахування ризиків маніпуляції даними чи результатами в політичних або соціальних цілях.

Отже, сучасні інновації та можливі вдосконалення методів ретропрогнозування дозволяють не лише підвищити точність і достовірність результатів, а й зробити їх доступними для міждисциплінарного аналізу. Інтеграція новітніх технологій, таких як AI, VR/AR і GIS, відкриває нові перспективи у вивченні минулого та його впливу на сучасність і майбутнє.

Цифровізація архівів і створення спеціалізованих баз даних відкривають нові перспективи для ретропрогнозування, забезпечуючи дослідників доступом до величезних обсягів

інформації та спрощуючи процес її аналізу. Ці можливості охоплюють як технічні, так і концептуальні аспекти роботи з історичними даними.

1. Оцифрування історичних документів.

Збереження історичних даних – старі рукописи, книги, карти, статистичні звіти та інші документи, які могли бути втрачені або зруйновані, тепер оцифровуються та зберігаються в цифровому форматі.

Покращення доступу – дослідники з будь-якого куточка світу можуть отримати доступ до архівних матеріалів через онлайн-бази даних. Це суттєво скорочує витрати на поїздки до архівів або бібліотек.

2. Автоматизовані інструменти для пошуку та аналізу.

Оптичне розпізнавання тексту (OCR) – технології OCR дозволяють швидко переводити рукописні або друковані тексти у машинозчитуваний формат, спрощуючи їхній аналіз.

Пошук за ключовими словами – інтеграція інтелектуальних алгоритмів у бази даних дозволяє здійснювати пошук не лише за ключовими словами, але й за концептуальними зв'язками.

3. Створення спеціалізованих баз даних.

Мультидисциплінарні дані – наприклад, бази даних, які поєднують демографічну, економічну, кліматичну й соціокультурну інформацію, дозволяють проводити глибокий аналіз історичних процесів.

Динамічні бази – системи, які регулярно оновлюються новими даними або доповнюються результатами сучасних досліджень, забезпечують актуальність інформації.

Глобальне охоплення – створення баз даних, які об'єднують архіви з різних країн, сприяє порівняльним дослідженням і аналізу взаємозв'язків між регіонами.

4. Інтеграція геопросторових даних.

Цифрові карти та геоінформаційні системи (GIS) – оцифровані історичні карти дозволяють аналізувати просторові зміни та їхній вплив на історичні процеси.

Просторово-часові бази даних – системи, які поєднують дані про місце та час подій, полегшують вивчення складних динамічних процесів, таких як міграція чи зміни клімату.

5. Відкриті платформи та співпраця.

Колаборація між науковцями – цифрові архіви та бази даних полегшують співпрацю між дослідниками різних дисциплін і країн.

Громадянська наука – доступність баз даних дозволяє залучати громадськість до досліджень, наприклад, для ідентифікації чи класифікації даних у великих проєктах.

Відкриті архіви – ініціативи з відкритого доступу забезпечують безкоштовний доступ до даних для освітніх і дослідницьких цілей.

6. Інноваційні технології аналізу даних.

Штучний інтелект і машинне навчання – алгоритми дозволяють автоматизувати аналіз великих масивів даних, виявляючи закономірності, які раніше могли залишатися непоміченими.

Семантичний аналіз – використання інструментів для аналізу контексту текстів дозволяє реконструювати ідеї та концепції минулого.

Попри значний потенціал, цифровізація також супроводжується викликами: а) дані з різних архівів можуть бути несумісними через різні формати або відсутність стандартизації; б) обробка великих обсягів даних вимагає значних обчислювальних потужностей; в) оцифровані матеріали іноді залишаються недоступними через правові обмеження.

Отже, цифровізація архівів і створення баз даних докорінно змінюють можливості ретропрогнозування, відкриваючи доступ до величезного обсягу інформації та сприяючи глибшому розумінню минулого. Удосконалення цих процесів забезпечує нові перспективи для міждисциплінарних досліджень, водночас сприяючи збереженню культурної спадщини для майбутніх поколінь.

Для ретропрогнозування використовують різноманітні імітаційно-прогностичні моделі, які є важливим інструментом для

дослідження складних систем і процесів. Вони дозволяють відтворювати реальні явища, аналізувати їх поведінку в умовах різних сценаріїв і робити прогнози. Типи таких моделей класифікуються залежно від підходу до моделювання, методу аналізу або сфери застосування.

1. *Дискретно-подієві моделі (Discrete Event Simulation, DES)* – моделі, що імітують системи, де зміни відбуваються у вигляді дискретних подій у певні моменти часу (наприклад, логістичні системи (моделювання черг, руху транспорту), управління виробничими процесами). Такі моделі використовуються для систем із чітко визначеними етапами або станами.

2. *Агентно-орієнтовані моделі (Agent-Based Models, ABM)* – моделі, у яких система складається з окремих агентів, що взаємодіють між собою відповідно до визначених правил (наприклад, соціальні моделі (взаємодія людей у суспільстві), економічні моделі (поведінка споживачів чи ринкових агентів)). Такі моделі підходять для систем, де поведінка частин суттєво впливає на загальну динаміку.

3. *Системно-динамічні моделі (System Dynamics, SD)* – моделі, що базуються на використанні диференціальних рівнянь для опису змін у системі як безперервного процесу (наприклад, екологічні системи (зміна популяції, ресурси), макроекономічні моделі (зростання економіки, інфляція)). Такі моделі застосовуються для аналізу тривалих змін у складних системах із багатьма взаємозалежними елементами.

4. *Стохастичні моделі* – моделі, які враховують випадковість і невизначеність у поведінці системи (наприклад, моделі ризиків (прогнозування природних катастроф), фінансові моделі (аналіз волатильності ринку)). Ці моделі включають ймовірнісні параметри, що відображають реальну непередбачуваність.

5. *Моделі на основі часових рядів (Time Series Models)* – моделі, що аналізують і прогнозують зміни змінних у часі на основі історичних даних (наприклад, кліматичні моделі (зміна температури),

економічні прогнози (ВВП, ціни)). Такі моделі використовують тренди, сезонність і залишкові компоненти для прогнозування.

6. *Геопросторові моделі (Geospatial Models)* – моделі, що враховують просторове розташування та взаємозв'язки між елементами системи (урбаністика (розвиток міст), прогнозування поширення епідемій). Такі моделі використовуються у поєднанні з геоінформаційними системами (GIS).

7. *Гібридні моделі* – моделі, що поєднують різні типи підходів (наприклад, агентно-орієнтовані та системно-динамічні). Наприклад, екологічні прогнози, що враховують поведінку окремих організмів і динаміку середовища; економіко-соціальні моделі з інтеграцією поведінкових і макроекономічних аспектів. Такі моделі забезпечують гнучкість і адаптивність при моделюванні складних систем.

8. *Моделі штучного інтелекту та машинного навчання* – моделі, які базуються на алгоритмах навчання з даних, використовуються для створення прогнозів на основі складних і багатовимірних залежностей. Наприклад, прогнозування попиту на ринку або аналіз поведінки споживачів. Такі моделі підходять для роботи з великими даними та для виявлення складних нелінійних залежностей.

9. *Кліматичні моделі* – моделі, які відтворюють і прогнозують зміни кліматичних умов на основі фізичних, хімічних і біологічних параметрів (наприклад, глобальні моделі кліматичних змін, прогнозування регіональних погодних умов). Ці моделі використовують фізичні закони для моделювання складних систем.

Типи імітаційно-прогностичних моделей різняться залежно від мети дослідження, доступних даних і специфіки системи, що моделюється. Кожен із цих типів має свої сильні сторони й обмеження, тому вибір моделі часто залежить від характеру задачі та доступних ресурсів. У багатьох випадках використання гібридного підходу дозволяє досягти більш точних і надійних результатів.

Крім наведеної класифікації часто операційно розрізняють альтернативні та контрфактичні імітаційно-прогностичні моделі для ретропрогнозування. Ці типи моделей є важливими інструментами для аналізу минулого, оскільки вони допомагають оцінити, як могли б розвиватися події за інших умов, і виявити ключові фактори, що впливали на історичні процеси.

Альтернативні імітаційно-прогностичні моделі дозволяють досліджувати варіанти розвитку подій, які могли б мати місце за зміни деяких умов або факторів. Ці моделі будуються на основі припущень про зміну одного або кількох параметрів системи, що не суперечить загальним законам або логіці реального світу. Метою застосування таких моделей є виявити, які можливості чи сценарії залишилися нереалізованими, і оцінити їх вплив на хід історичних подій. Наприклад: проведення аналізу, як би розвивалася економіка країни без певних історичних криз (наприклад, без Великої депресії); оцінка демографічних змін, якщо б у минулому було впроваджено інші політики, такі як контроль народжуваності.

Контрфактичні моделі будуються на припущеннях, які суперечать реальному розвитку подій. Вони дозволяють оцінити, як могли б розгортатися події, якщо ключові умови або події були б іншими. Ці моделі припускають умови, які фактично не існували, і досліджують, як би змінилися результати. Метою застосування таких моделей є розуміння причинно-наслідкових зв'язків, оцінка впливу конкретних подій або рішень, виявлення альтернативних шляхів розвитку. Наприклад, застосування таких моделей у політичній історії може дати можливість з'ясувати, що б сталося, якби конкретна битва була програна/виграна іншою стороною (наприклад, аналіз наслідків виграшу Наполеона при Ватерлоо). Або як би змінилася історія, якби певна технологія була винайдена раніше (наприклад, поява електрики в XVII столітті)? Або що було б, якби певні реформи були реалізовані в іншій формі або зовсім не були впроваджені?

Розглянемо ключові особливості та відмінності названих моделей у вигляді таблиці.

Характеристика	Альтернативні моделі	Контрфактичні моделі
Базові припущення	Зміни в межах реальних ймовірностей	Зміни, які суперечать реальному розвитку
Мета	Вивчення потенційно можливих сценаріїв	Аналіз причинно-наслідкових зв'язків
Приклад використання	Аналіз сценаріїв економічного зростання	Вивчення наслідків відсутності війни

До технологій реалізації таких моделей належать *методи моделювання* (системна динаміка для побудови сценаріїв, агентно-орієнтовані моделі для вивчення індивідуальних впливів), *штучний інтелект і машинне навчання* (аналіз великих історичних даних і побудова сценаріїв із використанням нейронних мереж, інтеграція контрфактичного аналізу у прогностичні моделі), *симуляції* (моделювання альтернативних історичних реалій у віртуальному середовищі).

При використанні таких моделей існують такі виклики та обмеження: *невизначеність даних* (для створення моделей необхідні якісні та детальні історичні дані, які не завжди доступні), *складність верифікації* (неможливо однозначно перевірити точність альтернативних або контрфактичних сценаріїв), *суб'єктивність* (моделі значною мірою залежать від припущень і вибору змінних дослідниками).

Альтернативні та контрфактичні моделі мають важливе значення для ретропрогнозування, так як дозволяють *розкрити ключові фактори минулого* (допомагають зрозуміти, які події чи рішення найбільше вплинули на розвиток історії), *підтримувати прийняття рішень у майбутньому* (аналіз історичних сценаріїв може бути використаний для розробки стратегій уникнення криз або повторення успішних сценаріїв), *покращити розуміння історичних*

процесів (сприяють інтеграції різних дисциплін, таких як економіка, соціологія, історія, для комплексного аналізу минулого).

Отже, альтернативні та контрфактичні моделі є потужними інструментами, що розширюють можливості ретропрогнозування, дозволяючи дослідникам аналізувати не лише те, що відбулося, але й те, що могло б статися за інших умов. Це відкриває нові перспективи в розумінні історії та створює базу для інформованих рішень у майбутньому.

Досить часто при ретропрогнозування застосовується побудова динамічних рядів за неповними даними, яка є важливим аспектом дослідження, оскільки історичні записи часто бувають фрагментарними, неточними або неповними. Основне завдання – відновлення втрачених або відсутніх даних та створення узгодженого ряду для аналізу минулих тенденцій і прогнозування. Побудова таких динамічних рядів відбувається у декілька етапів.

1. *Проводиться аналіз наявних даних* – визначається якість, обсяг та достовірність доступних історичних даних, виявляються відсутні періоди чи елементи у часовому ряді, проводиться класифікація даних (дані можуть бути представлені як кількісні (цифрові) або якісні (описові), що впливає на підхід до заповнення).

2. *Проводиться попередня обробка даних* – дані очищаються (усунення аномалій, помилок введення, дублювання та інших спотворень), проводиться їх нормалізація (приведення даних до єдиної шкали чи формату для полегшення обчислень).

3. *Відновлюються відсутні значення* – реконструкція пропущених даних здійснюється через застосування методів і підходів, які враховують особливості системи: а) дотримання принципу інтерполяції – заповнення відсутніх значень шляхом припущення, що зміни між наявними точками є лінійними або слідує певній залежності; б) застосування методів лінійної інтерполяції, поліноміальної інтерполяції, кубічних сплайнів; в) застосування методу екстраполяції – прогнозування відсутніх значень на основі наявного тренду даних, або використання для реконструкції віддалених точок, коли доступні лише неповні частини ряду;

г) застосування статистичних моделей – заповнення пропущених точок середнім значенням наявних даних, заповнення пропущених точок медіанним значенням наявних даних (особливо корисне для рядів із сильними аномаліями чи викидами); д) застосування методів машинного навчання – використання алгоритмів, які навчаються на доступних даних і прогнозують пропущені точки (наприклад, логістична регресія, глибокі нейронні мережі, рекурентні нейронні мережі (RNN)); е) застосування моделей часових рядів – використання спеціалізованих методів аналізу часових рядів, таких як ARIMA (AutoRegressive Integrated Moving Average) для прогнозування трендів і циклів, SARIMA (Seasonal ARIMA) для врахування сезонних коливань; є) застосування контекстуальних моделей – залучення додаткових джерел даних (географічних, економічних, соціальних) для уточнення відсутніх значень.

4. *Проводиться верифікація та узгодження* – перевіряється логічна послідовність (відновлені дані мають відповідати відомим історичним закономірностям чи подіям), проводиться перевірка трендів (оцінка узгодженості трендів після заповнення прогалін), організовується тестування моделей (використання метрик, таких як середньоквадратична помилка (RMSE) або середня абсолютна помилка (MAE), для перевірки точності відновлення).

5. *Проводиться візуалізація та аналіз даних часового ряду* – будується графічне представлення (побудова графіків (діаграм, гістограм і т.п.) для візуальної оцінки відновленого ряду), проводиться порівняння (аналіз відновлених даних із наявними зовнішніми джерелами чи альтернативними моделями).

Інструментами для обробки неповних даних часових рядів можуть виступати різноманітне програмне забезпечення (наприклад, Python – бібліотеки pandas, NumPy, statsmodels; R – пакети zoo, forecast, Amelia; MATLAB – функції для аналізу часових рядів) та спеціалізовані програми (SPSS, STATA для статистичного аналізу, Tableau, Power BI для візуалізації та аналізу динамічних рядів).

При побудові динамічних рядів з неповними даними можуть виникати такі виклики: *нестача якісних даних* (історичні дані часто містять помилки, спотворення або є неповними), *складність моделювання* (часові ряди можуть мати нелінійні залежності, які важко точно відтворити), *проблеми узгодженості* (відновлені дані повинні відповідати відомим історичним фактам), *суб'єктивність* (вибір методів відновлення залежить від дослідника, що може впливати на результат).

Загалом, побудова динамічних рядів за неповними даними дозволяє відновити історичні тренди, необхідні для аналізу та моделювання, поліпшити точність ретропрогнозів через узгодження минулих подій із сучасними моделями, інтегрувати різні джерела даних для створення комплексних моделей минулого. Цей процес є основою для якісного аналізу та прогнозування у багатьох сферах, включаючи економіку, кліматологію, соціальні науки та археологію.

Перспективними напрямками покращення точності ретропрогнозів є використання штучного інтелекту та машинного навчання. Штучний інтелект (ШІ) та машинне навчання (МН) є ключовими технологіями, які сприяють значному вдосконаленню ретропрогнозування завдяки їхній здатності обробляти великі обсяги даних, виявляти приховані закономірності та адаптуватися до нових типів даних. Ці технології відкривають нові можливості для підвищення точності ретропрогнозів і їхнього застосування в різних галузях.

По-перше, при обробці великих обсягів даних ШІ здатний обробляти масиви даних, що походять із різних джерел (текстів, зображень, картографічних даних, числових показників) значно швидше й ефективніше, ніж традиційні методи. При цьому МН може поєднувати числові, текстові, візуальні та геопросторові дані для створення більш комплексних моделей.

По-друге, при виявленні закономірностей та трендів алгоритми МН можуть знаходити закономірності, які неочевидні для дослідників, що особливо корисно для аналізу складних історичних

процесів. ШІ дозволяє виявляти події чи процеси, які не відповідають загальним тенденціям, що допомагає уточнювати історичні моделі.

По-третє, при використанні глибокого навчання (Deep Learning) можливою є обробка неструктурованих даних (наприклад, глибокі нейронні мережі можуть аналізувати стародавні тексти, рукописи чи археологічні записи, автоматично "підтягуючи" відповідні дані для прогнозів) та застосування нейронних мереж для часових рядів (моделі, такі як LSTM (Long Short-Term Memory) та трансформери, дозволяють аналізувати часові ряди для відтворення складних змін у часі).

По-четверте, розвиток байєсівських моделей дає можливість проведення аналізу із невизначеностями (байєсівські підходи дозволяють враховувати нестачу або неточність даних, створюючи ймовірнісні прогнози) та організації процедур оновлення моделей (ШІ може динамічно оновлювати ретропрогнози за появи нових даних, підвищуючи їхню точність).

По-п'яте, технології МН можуть покращувати реконструкцію даних – алгоритми МН можуть заповнювати прогалини у даних, відновлюючи втрачені або пошкоджені фрагменти текстів, зображень або часових рядів. Глибоке навчання дозволяє «оживити» старі карти, визначити географічні зміни та впливати на їхнє точніше трактування.

По-шосте, за допомогою ШІ можна будувати персоналізовані та адаптивні моделі – моделі ШІ здатні вдосконалюватися в процесі роботи, адаптуючись до специфіки конкретних наборів даних. Алгоритми можуть бути налаштовані під конкретні дослідницькі задачі, такі як аналіз економічних криз, демографічних змін чи соціальних трансформацій.

По-сьоме, інтеграція ШІ з іншими інноваційними технологіями, наприклад, з GIS-системами, дозволяє аналізувати просторові закономірності й виявляти взаємозв'язки між географічними, кліматичними та соціальними даними.

Використання ШІ можливе у віртуальних моделях для тестування альтернативних сценаріїв історичних подій.

По-восьме, технології ШІ та МН дадуть можливість полегшення візуалізації та інтерпретації результатів – інструменти ШІ створюють інтерактивні графіки та 3D-моделі, які полегшують сприйняття результатів для дослідників і громадськості. Алгоритми інтерпретації дозволяють пояснювати результати навіть для користувачів без глибоких технічних знань.

Попри значний потенціал, використання ШІ у ретропрогнозуванні стикається з низкою викликів: а) якість даних – алгоритми залежать від точності й повноти історичних записів, а низька якість може призводити до спотворених результатів; б) пояснюваність моделей – складні алгоритми глибокого навчання часто є «чорними скриньками», що ускладнює їхню інтерпретацію; в) етичні аспекти – існує ризик маніпуляції даними або результатами для досягнення політичних чи ідеологічних цілей.

Таким чином, інтеграція штучного інтелекту та машинного навчання у ретропрогнозування відкриває нові горизонти для дослідників. Завдяки цим технологіям можна створювати більш точні, адаптивні й деталізовані моделі, що дозволяють краще зрозуміти минуле. Водночас подальший розвиток цих підходів потребує не лише технічного вдосконалення, а й уваги до етичних і методологічних аспектів.

7.2. Застосування інструментів EXCEL для ретропрогнозування

Прогнозування – хоч і невдячна, але необхідна справа. Для вирішення таких завдань у Microsoft Excel є дуже пристойний інструментарій – від найпростіших функцій лінійного тренду (див. розділ 5 посібника) до крутих статистичних інструментів з надбудови *Пакет Аналізу (Analysis Toolpak)*. Одними з найпростіших

у реалізації і при цьому дуже ефективними є функції прогнозування методом експоненційного згладжування.

Суть цього (якщо не вдаватися в математичні подробиці) можна пояснити відносно легко. Якби ми, наприклад, робили прогноз дуже примітивним способом за середнім арифметичним, то всі історичні дані бралися б з однаковою вагою (у статистиці цей метод «середньої температури по лікарні» має, до речі, навіть офіційну назву – «наївний прогноз»). При прогнозуванні ж методом експоненційного згладжування приймається ідея, що старі дані повинні мати вагу менше, ніж нові. Зміна цієї ваги в залежності від новизни або старості наших даних відбувається за лавиноподібною кривою експоненціальної – звідси і назва методики. Застосування методу експоненційного згладжування у надбудові *Пакет Аналізу Microsoft Excel* ми показали у п. 5.2 посібника.

У Microsoft Excel для реалізації операції прогнозування є дві основні функції, що з'явилися, починаючи з версії Excel 2016 року:

ПЕРЕДКАЗ.ETS (FORECAST.ETS) – обчислює майбутні спрогнозовані значення на основі історичних даних.

ПЕРЕДКАЗА.ETS.ДОВІНТЕРВАЛ (FORECAST.ETS.CONFINT) – обчислює розмах довірчого інтервалу – коридору похибки, в межах якого із заданою ймовірністю наш прогноз має здійснитися.

Втім, вводити ці функції і їх численні аргументи вручну зовсім не потрібно – в Microsoft Excel для цього є набагато більш зручний інструмент, що отримав назву *Лист прогнозу (Forecast Sheet)*. Розглянемо роботу з ним на наступному прикладі.

Нехай у нас є масив динамічних даних, що представляє собою кількісні значення такого явища: динаміка зміни середнього значення забезпеченості закладів вищої освіти України персональними комп'ютерами навчального призначення у розрахунку на 100 студентів денної форми навчання протягом 1991–2011 рр.

Рік	Значення
1991	7,83
1992	8,60
1993	9,38
1994	9,19
1995	8,31
1996	8,45
1997	10,81
1998	12,10
1999	14,29
2000	16,90
2001	20,64
2002	20,45
2003	22,57
2004	23,86
2005	25,34
2006	26,16
2007	26,71
2008	28,56
2009	28,98
2010	29,14
2011	30,25

Спрогнозуємо таку динаміку до 2030 року. Виділимо всю нашу таблицю і на вкладці *Дані* в групі *Прогноз* скористаємося кнопкою *Аркуш прогнозу (Data – Forecast Sheet)*. У вікні, що відкрилося, задаємо наступні налаштування:

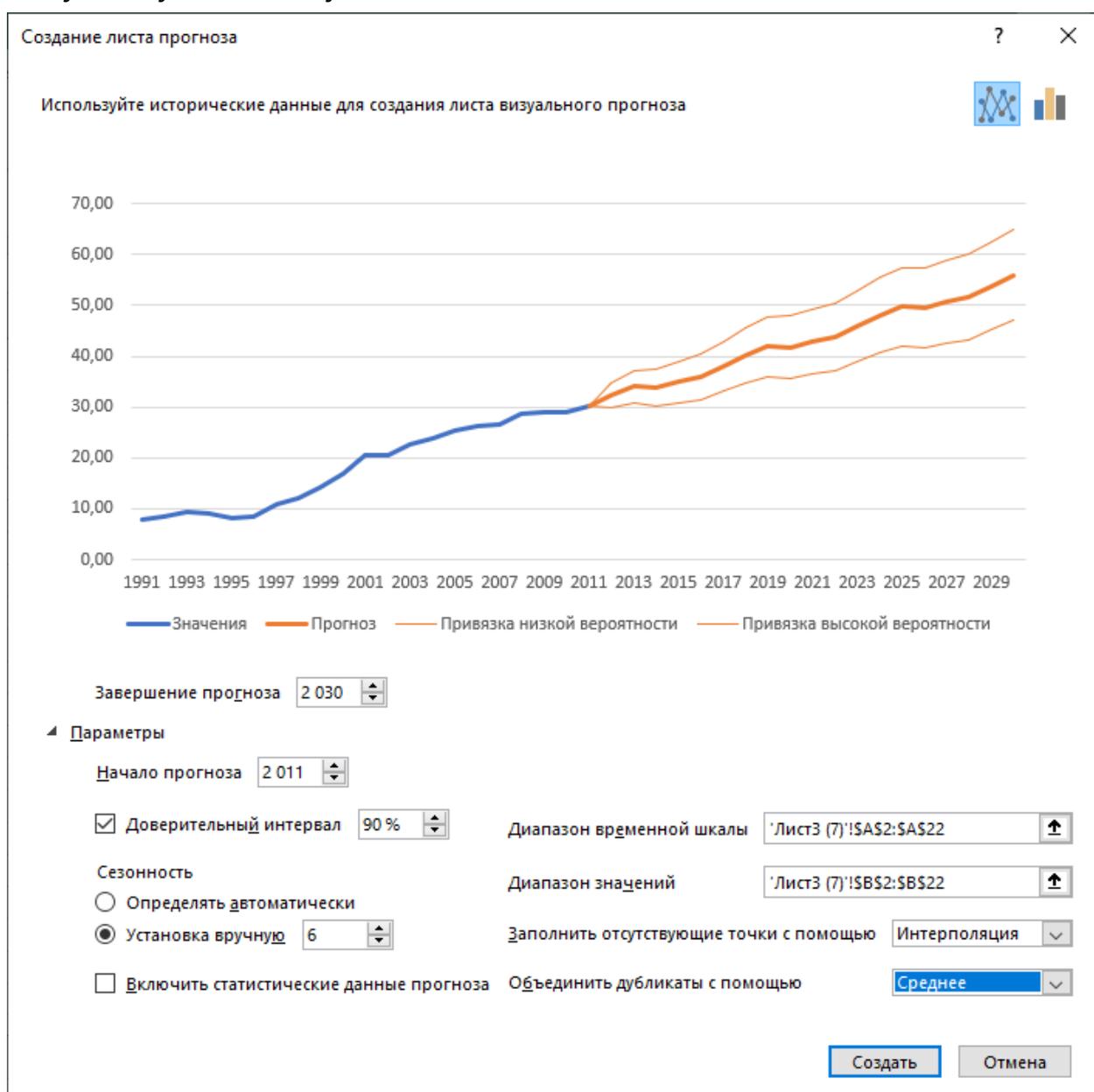
Дату завершення прогнозу – вказуємо 2030 рік.

Сезонність – майже ніколи коректно не визначається автоматично на жаль, тож краще задати її вручну. Тому задамо сезонність 6, гіпотетично враховуючи період, протягом якого у ЗВО України двічі замінюється відповідне технічне обладнання.

Ймовірність, з якою ми вимагаємо попадання майбутніх фактичних значень до коридору довірчого інтервалу. Чим більша ця ймовірність, тим ширший інтервал (тобто більш розмитий

прогноз). Зазвичай використовують значення 90–95%. Обираємо значення 90%.

У правому нижньому кутку вікна можна додатково вибрати реакцію на *порожні комірки* (їх можна заповнити нулями або середнім сусіднім значенням – інтерполяцією) та на *дублікати* (зазвичай їх усереднюють). Однак, по можливості, краще заздалегідь підготувати вихідні історичні дані, щоб таких прогалів чи дублів у них не було.



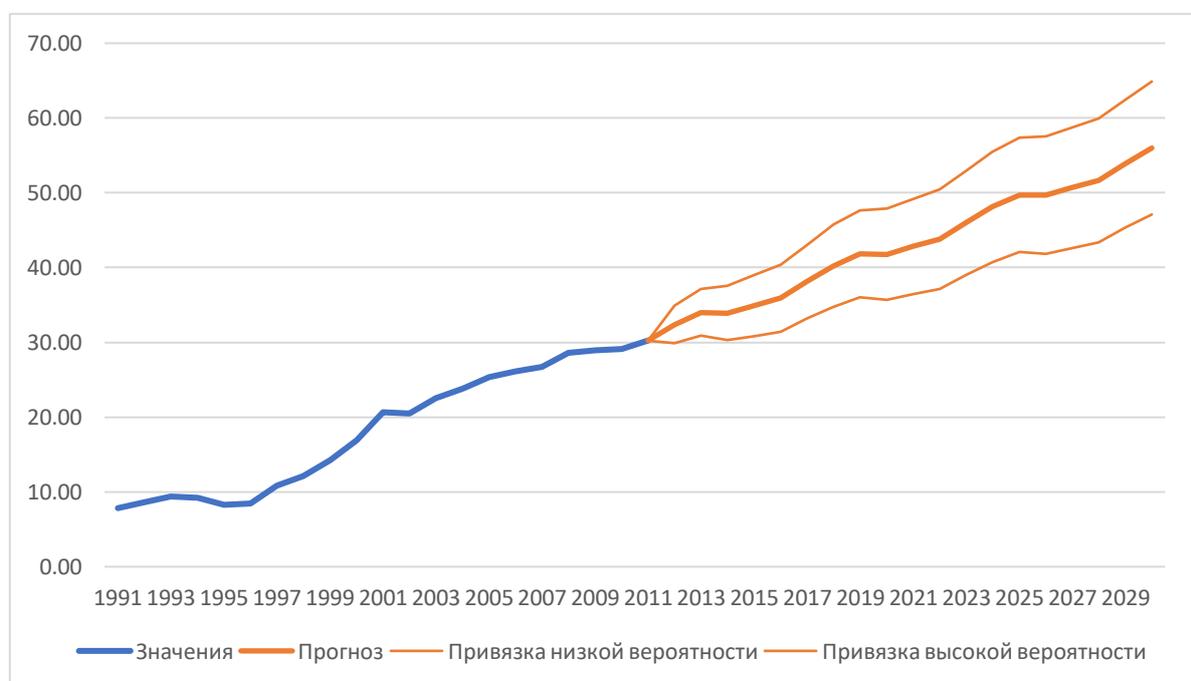
У діалоговому вікні *Створення аркуша прогнозу* виберемо лінійчасту або стовпчасту діаграму, щоб наочно представити

прогноз. Після натискання на кнопку *Створити* Excel створює новий аркуш із таблицею минулих і прогнозованих значень. Новий аркуш можна знайти ліворуч аркуша, на якому ми вводили ряди даних.

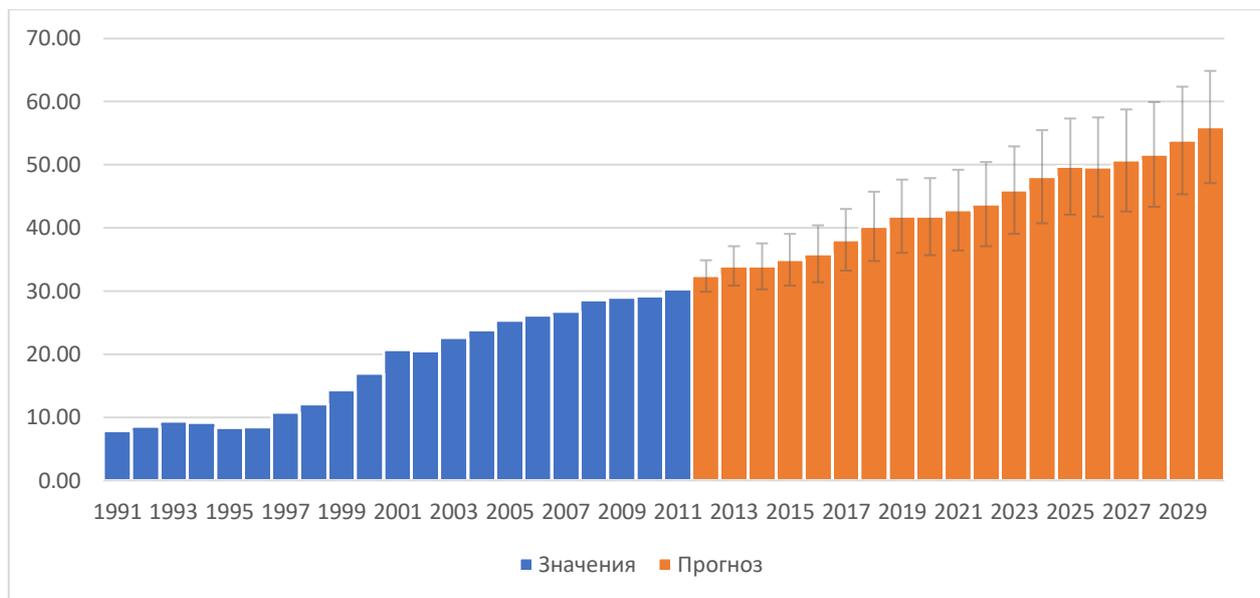
Временная шкала	Значения	Прогноз	Привязка низкой вероятности	Привязка высокой вероятности
1991	7,83			
1992	8,60			
1993	9,38			
1994	9,19			
1995	8,31			
1996	8,45			
1997	10,81			
1998	12,10			
1999	14,29			
2000	16,90			
2001	20,64			
2002	20,45			
2003	22,57			
2004	23,86			
2005	25,34			
2006	26,16			
2007	26,71			
2008	28,56			
2009	28,98			
2010	29,14			
2011	30,25	30,25	30,25	30,25
2012		32,38	29,89	34,88
2013		33,99	30,87	37,10
2014		33,92	30,28	37,56
2015		34,96	30,86	39,05
2016		35,90	31,40	40,41

2017	38,12	33,24	43,00
2018	40,25	34,76	45,74
2019	41,85	36,05	47,66
2020	41,79	35,68	47,89
2021	42,82	36,43	49,22
2022	43,77	37,10	50,44
2023	45,99	39,06	52,92
2024	48,12	40,74	55,50
2025	49,72	42,10	57,34
2026	49,65	41,80	57,51
2027	50,69	42,61	58,77
2028	51,64	43,34	59,94
2029	53,86	45,33	62,38
2030	55,99	47,10	64,88

Можна також вивести лінійчасту діаграму, на якій буде представлено значення з таблиці минулих і прогнозованих значень:



Крім такої діаграми можна вивести і стовпчасту діаграму.



Дамо кілька рекомендацій щодо налаштування прогнозів. Щоб змінити будь-які додаткові параметри прогнозу, клацніть елемент *Параметри*.

1. *Початок прогнозу*. Виберіть дату початку прогнозу. Якщо вибрати дату, що передує останній даті в даних за минулий період, для прогнозу буде використано лише дані до дати початку (такий прогноз називається ретроспективним). Прогноз, що починається перед останньою точністю в даних за минулий період, дає можливість відчутти точність прогнозування, порівняти прогнозований ряд із фактичними даними. Проте якщо почати прогноз зарано, створений прогноз не обов'язково збігатиметься з тим, який можна отримати, використавши всі дані за минулий період. Якщо використати всі минулі дані, прогноз буде точніший. Якщо дані сезонні, радимо починати прогноз із дати, яка передує останній даті в даних за минулий період.

2. *Довірчий інтервал*. Установіть або зніміть прапорець Довірчий інтервал, щоб відобразити або приховати цей інтервал. Довірчий інтервал – це діапазон навколо кожного прогнозованого значення, у який, відповідно до прогнозу, має потрапити 95% майбутніх точок (у випадку нормального розподілу). За допомогою довірчого інтервалу можна краще оцінити точність прогнозу. Менший інтервал свідчить про більшу достовірність прогнозу для

конкретної точки. Стандартне значення довірчого рівня – 95%, але його можна змінити за допомогою стрілок угору або вниз.

3. *Сезонний фактор*. Сезонний характер – це число для довжини (кількості точок) сезонного шаблону й виявляється автоматично. Наприклад, у річному циклі кожна точка представляє місяць, сезонність – 12. Автоматичне виявлення можна змінити, вибравши Установити вручну, а потім вибравши число. Настроюючи сезонність вручну, використовуйте дані за період, що охоплює не менше 2 циклів. Якщо циклів менше 2, програма Excel не може визначити сезонні компоненти. Якщо сезонність недостатньо значуща, щоб її міг виявити алгоритм, для прогнозу використовуватиметься лінійне наближення.

4. *Діапазон часової шкали*. Тут можна змінити діапазон, який використовуватиметься для часової шкали. Цей діапазон має збігатися зі значенням поля *Діапазон значень*.

5. *Діапазон значень*. Тут можна змінити діапазон, який використовуватиметься для ряду значень. Цей діапазон має збігатися зі значенням поля *Діапазон часової шкали*.

6. *Заповнення відсутніх точок*. Якщо деякі точки відсутні, Excel застосовує метод інтерполяції, тобто замість значень відсутніх точок використовується зважене середнє значення сусідніх точок (якщо бракує менше 30% точок). Щоб замість значень відсутніх точок використовувалися нулі, виберіть зі списку значення Нулі.

7. *Агрегатні повторення за допомогою*. Якщо дані містять кілька значень з однаковою позначкою часу, Excel використовує середні значення. Щоб скористатися іншим методом обчислення, наприклад *Медіана* або *Кількість*, виберіть потрібний метод у списку.

8. *Включити статистичні дані прогнозування*. Установіть цей прапорець, якщо потрібно включити додаткові статистичні відомості до нового аркуша. Таким чином буде створено таблицю статистики з використанням функції FORECAST.ETS. Функція STAT містить такі міри, як коефіцієнти згладжування (альфа, бета, гама) і показники помилок (MASE, SMAPE, MAE, RMSE).

РОЗДІЛ 8.

МЕТОДИ БАГАТОМІРНОГО СТАТИСТИЧНОГО АНАЛІЗУ В ІСТОРИЧНИХ ДОСЛІДЖЕННЯХ

8.1. Принципи багатомірного статистичного аналізу. Кластерний, дискримінантний, факторний аналіз. Багатомірна класифікація

Сучасна історична наука дедалі більше інтегрує міждисциплінарні методи, які дозволяють не лише інтерпретувати джерела, а й кількісно аналізувати великі масиви даних. У цьому контексті методи багатомірного статистичного аналізу (БСА) стають важливим інструментом для вивчення складних історичних явищ, що мають багатовимірний характер. Завдяки своїй здатності опрацьовувати різнорідні змінні, БСА дозволяють побудувати комплексну картину минулого, враховуючи економічні, соціальні, демографічні та культурні аспекти.

Доцільність використання багатомірного статистичного аналізу в історичних дослідженнях пояснюється зростанням обсягів доступних даних, зокрема завдяки оцифруванню архівів, реєстрів та історичних текстів. Методи БСА дозволяють систематизувати ці дані, виявляти приховані закономірності та проводити узагальнення, які неможливо отримати за допомогою традиційних підходів.

Метою цього розділу є висвітлення основних методів багатомірного статистичного аналізу, їхньої сутності, особливостей застосування в історичних дослідженнях та інтерпретації результатів. Основна увага приділяється теоретичним засадам методів, практичним аспектам їх використання та прикладам з історичної науки.

У межах розділу розглядаються такі ключові аспекти: огляд основних методів багатомірного аналізу, особливості їхньої адаптації до історичних даних, етапи проведення аналізу та критичний огляд переваг і обмежень. Крім того, окреслюються перспективи подальшого впровадження цих методів у міждисциплінарних історичних дослідженнях.

Цей підхід дозволяє не лише переосмислити традиційні питання історичної науки, але й підвищити її методологічну базу, відкриваючи нові можливості для розуміння минулого.

Актуальність використання БСА в історичних дослідженнях зумовлена кількома ключовими чинниками.

По-перше, історична наука сьогодні має доступ до великих масивів даних завдяки оцифруванню архівних документів, реєстрів, переписів населення, економічних звітів тощо. Традиційні методи аналізу не завжди дозволяють ефективно опрацьовувати таку кількість даних, тоді як БСА забезпечує систематизацію та комплексну обробку великих інформаційних масивів.

По-друге, історичні явища часто мають багатовимірний характер, адже на їх розвиток впливають численні соціальні, економічні, політичні, культурні й географічні чинники. Методи БСА дозволяють враховувати ці численні фактори одночасно, визначати їх взаємозв'язки та виділяти найважливіші з них.

По-третє, БСА дозволяє знаходити закономірності, які неможливо ідентифікувати за допомогою традиційного описового аналізу. Наприклад, використання кластерного аналізу допомагає групувати явища за схожістю, а факторний аналіз дає змогу виявити приховані структури у великих масивах історичних даних.

По-четверте, інтеграція методів із суміжних галузей, таких як соціологія, економіка та інформатика, сприяє більшій об'єктивності та глибині історичних висновків. БСА надає історикам інструменти, які давно використовуються в інших науках, що дозволяє поєднувати якісний і кількісний підходи.

По-п'яте, БСА допомагає аналізувати довгострокові тенденції, наприклад, демографічні зміни, економічні кризи чи міграційні процеси. Завдяки цьому історики можуть будувати математично обґрунтовані моделі, що пояснюють минулі події.

По-шосте, застосування математичних методів зменшує вплив суб'єктивізму дослідника на процес інтерпретації даних. Це дозволяє отримувати більш точні та обґрунтовані результати, що є особливо важливим у дискусійних питаннях історії.

Нарешті, БСА відкриває нові горизонти для історичної науки. Наприклад, аналіз текстових корпусів (письмових джерел) за допомогою методів машинного навчання та статистики, а також дослідження великих даних у контексті цифрової історії.

Таким чином, багатомірний статистичний аналіз є незамінним інструментом у сучасних історичних дослідженнях. Його використання сприяє якіснішому аналізу джерел, відкриттю нових закономірностей та формуванню комплексного уявлення про історичні явища.

Багатомірний статистичний аналіз – це сукупність математичних і статистичних методів, які використовуються для дослідження й інтерпретації даних, що містять кілька взаємозалежних змінних. Його основна мета полягає у виявленні закономірностей, залежностей і структур у багатовимірних наборах даних, а також у спрощенні їх для подальшої інтерпретації.

Основні особливості реалізації БСА є такими:

1. *Робота з багатьма змінними* – аналіз проводиться одночасно для кількох змінних, що дозволяє враховувати їх взаємозв'язок та вплив на досліджуване явище.

2. *Зменшення розмірності* – часто дані містять зайву або повторювану інформацію. БСА дозволяє зменшити розмірність

набору даних (наприклад, за допомогою аналізу головних компонент), зберігаючи основну інформацію.

3. *Виявлення прихованих структур* – БСА дозволяє виявляти приховані закономірності або групи (кластеризація), які не є очевидними при поверхневому аналізі.

Завдяки методам БСА дослідники можуть будувати узагальнені моделі, що відображають основні риси досліджуваного явища.

Основними компонентами БСА є такі:

1. *Дані* – дані для БСА містять кілька змінних, що характеризують об'єкти дослідження. Важливо забезпечити стандартизацію даних, оскільки різні одиниці виміру можуть спотворити результати.

2. *Методи* – методи БСА включають кластерний аналіз, факторний аналіз, аналіз головних компонент, багатовимірне шкалювання, дискримінантний аналіз тощо.

3. *Результати* – результатом є графічні або числові моделі, що пояснюють взаємозв'язки між змінними чи групами об'єктів.

Отже, сутність БСА полягає у спробі пояснити складні системи даних через їхнє структурування, упорядкування та виявлення взаємозв'язків.

БСА базується на таких ключових завданнях:

1. *Дослідження взаємозв'язків* – аналіз залежності між багатьма змінними одночасно, наприклад, між економічними, демографічними й соціальними факторами в історичних дослідженнях.

2. *Класифікація* – поділ об'єктів на групи за певними критеріями. Наприклад, кластеризація міст за рівнем економічного розвитку в різні історичні періоди.

3. *Скорочення розмірності* – спрощення складних багатовимірних моделей для їхньої легшої інтерпретації, наприклад, через виділення ключових факторів, які впливають на певне історичне явище.

4. *Побудова моделей* – використання даних для побудови моделей, що пояснюють історичні явища або прогнозують їх подальший розвиток.

Таким чином, багатомірний статистичний аналіз є потужним інструментом для дослідження складних багатовимірних явищ. Його застосування в історії дозволяє систематизувати великі обсяги даних, виділяти найважливіші закономірності та будувати обґрунтовані моделі для пояснення минулих процесів.

Методи багатомірного статистичного аналізу мають багатовікову історію, яка відображає поступовий розвиток математичних і статистичних підходів для роботи з великими й складними наборами даних. Основні етапи становлення БСА можна поділити на кілька ключових періодів:

1. *Ранній етап (XVIII–XIX століття)* – виникнення базових концепцій статистики. Однією з перших спроб створити аналітичний підхід до багатовимірних даних була робота Йоганна Ламберта (XVIII століття), який використовував математичні методи для вивчення фізичних явищ. Карл Фрідріх Гаусс (XVIII століття) розробив метод найменших квадратів, який став основою для регресійного аналізу, що є базовим компонентом багатомірного аналізу. Адольф Кетле (XIX століття) – бельгійський статистик, який застосував статистику до соціальних явищ і запропонував поняття «середньої людини». Його роботи стали основою для соціометрії та демографії.

2. *Формування основ багатомірного аналізу (початок XX століття)*. Карл Пірсон (1901) заснував аналіз головних компонент (PCA), який використовується для зменшення розмірності даних і виявлення найважливіших змінних. Його методика стала ключовою для сучасного БСА. Рональд Фішер (1920-ті роки) розробив дискримінантний аналіз, який дозволяє класифікувати об'єкти на основі кількох змінних. Гарольд Готеллінг (1930-ті роки) удосконалив методи аналізу головних компонент і запропонував канонічний кореляційний аналіз для вивчення залежностей між

двома наборами змінних. Джон Тьюкі зробив значний внесок у розвиток описової статистики й візуалізації багатовимірних даних.

3. *Зрілість методів БСА (середина ХХ століття).* В цей період набув розвитку кластерний аналіз, з'явилися методи кластеризації, зокрема ієрархічний кластерний аналіз (Гудман і Крускал, 1950-ті роки), що використовуються для групування об'єктів. Набув поширення факторний аналіз (1940–1950-ті роки) завдяки роботам математиків і психологів, які використовували його для вивчення латентних змінних у соціології та психології. Розпочалося застосування мережевого аналізу (1950–1960-ті роки), розробленого для вивчення структурних зв'язків між елементами у великих системах, зокрема в соціальних мережах.

4. *Інтеграція з комп'ютерними технологіями (друга половина ХХ століття).* Розвиток комп'ютерів у 1960–1970-х роках значно прискорив виконання складних математичних операцій і зробив БСА доступнішим для дослідників. Поява спеціалізованих програм (SPSS, SAS, MATLAB) дозволила використовувати БСА у різних галузях науки, включаючи історію. Методи багатомірного аналізу почали широко застосовуватися в демографії, економіці, історії та інших дисциплінах для аналізу великих обсягів даних.

5. *Сучасний етап (ХХІ століття).* Відбуваються процеси інтеграції з машинним навчанням – сучасні методи, такі як кластеризація та аналіз головних компонент, активно використовуються в алгоритмах машинного навчання для обробки великих обсягів даних. Відбувся великий прорив у візуалізації даних – розроблені інструменти для візуалізації багатовимірних структур, таких як діаграми, теплові карти, мережеві графіки. Методи БСА почали застосовуватися в цифровій історії – аналіз текстових корпусів, оцифрування архівів та історичних баз даних за допомогою БСА дозволив проводити новаторські міждисциплінарні дослідження. Розвивається багатовимірна статистика в GIS-системах – географічні інформаційні системи (GIS) інтегрують методи БСА для аналізу просторово-часових історичних даних.

Історія розвитку БСА демонструє еволюцію від базових методів до сучасних комп'ютеризованих технологій, які дозволяють працювати з великими обсягами даних. Цей поступ зробив багатовимірний аналіз одним із найважливіших інструментів у багатьох дисциплінах, зокрема й у дослідженнях історичних явищ.

Багатомірний статистичний аналіз базується на низці фундаментальних підходів і принципів, які забезпечують ефективну роботу з багатовимірними даними. Ці принципи дозволяють дослідникам зосередитися на найважливішій інформації в складних наборах даних, виявляти приховані закономірності та будувати науково обґрунтовані висновки.

Ключовими підходами БСА є такі:

1. *Врахування багатовимірності явищ* – основна ідея полягає у врахуванні множини змінних, які можуть бути взаємозалежними. Усі змінні розглядаються в комплексі, що дозволяє виявляти їх взаємозв'язки та спільний вплив на досліджуване явище.

2. *Зменшення розмірності даних* – у багатовимірних даних часто присутня надмірність через корельованість змінних. Методи, такі як аналіз головних компонент (РСА), виділяють найсуттєвіші фактори, скорочуючи розмірність і зберігаючи важливу інформацію.

3. *Класифікація та групування* – один із підходів БСА спрямований на поділ об'єктів на групи за схожістю ознак. Кластерний аналіз, наприклад, дозволяє знаходити природні групи об'єктів у даних.

4. *Візуалізація багатовимірних структур* – БСА часто передбачає використання методів для візуального представлення багатовимірних даних (графіки, багатовимірні шкали, діаграми). Це полегшує інтерпретацію складних взаємозв'язків між змінними.

5. *Виявлення латентних змінних* – у багатьох випадках явища описуються прихованими факторами, які неможливо виміряти безпосередньо. Факторний аналіз і структурне моделювання допомагають визначити ці латентні змінні.

Багатовимірний статистичний аналіз базується на таких принципах:

1. *Цілісність підходу* – аналіз даних повинен розглядатися як єдиний процес: від збору даних до їхньої обробки, аналізу й інтерпретації. Усі етапи при цьому повинні бути логічно узгодженими.

2. *Стандартизація змінних* – для коректного порівняння змінних, виміряних у різних одиницях, їх стандартизують (перетворюють у показники з середнім значенням 0 і стандартним відхиленням 1). Це забезпечує адекватність результатів аналізу.

3. *Оптимізація співвідношення сигналу та шуму* – БСА прагне зменшити «шум» у даних, тобто випадкові чи незначущі коливання, і зосередитись на суттєвих закономірностях.

4. *Придатність моделей* – використовувані методи мають відповідати характеру даних та дослідницьким завданням. Наприклад, для номінальних змінних застосовуються одні методи, а для кількісних – інші.

5. *Інтерпретація результатів* – результати багатомірного аналізу повинні мати зрозуміле й практичне тлумачення, яке відповідає контексту дослідження.

6. *Репрезентативність даних* – якість багатомірного аналізу значною мірою залежить від того, наскільки репрезентативними є дані. Дані мають відповідати умовам дослідження, а вибірка – адекватно відображати генеральну сукупність.

7. *Адаптивність підходів* – методика аналізу може змінюватися залежно від характеру даних і мети дослідження. Зокрема, в історичних дослідженнях можуть адаптуватися методи для роботи з неповними або різномірними джерелами.

Отже, ключові підходи та принципи багатомірного статистичного аналізу орієнтовані на комплексну обробку та інтерпретацію даних. Вони забезпечують гнучкість і точність під час аналізу багатовимірних явищ, зберігаючи при цьому чітку логіку дослідження. У застосуванні до історії ці підходи дозволяють

інтегрувати численні джерела інформації, знаходити приховані зв'язки між подіями та робити ґрунтовні наукові висновки.

Багатомірний статистичний аналіз включає широкий спектр методів, які застосовуються для аналізу даних, що містять кілька змінних одночасно. Методи БСА можна умовно розділити на три основні групи: методи зменшення розмірності, методи класифікації та методи виявлення зв'язків.

1. Методи класифікації та групування.

Ці методи спрямовані на поділ об'єктів на класи або групи за схожістю. Серед таких методів називають: а) *кластерний аналіз – групування об'єктів у кластери так, щоб об'єкти в одному кластері були подібними; при цьому використовуються ієрархічний кластерний аналіз, метод k-середніх (k-means), метод DBSCAN (кластеризація на основі щільності); б) дискримінантний аналіз – класифікація об'єктів за заздалегідь визначеними групами, за допомогою якої визначають, які змінні найкраще розрізняють групи; в) регресійні методи для багатомірних даних – проводиться для побудови моделей, що пояснюють залежність однієї змінної від кількох предикторів, з використанням лінійної багатовимірної регресії та логістичної регресії (для якісних залежних змінних).*

2. Методи зменшення розмірності.

Ці методи використовуються для спрощення складних багатовимірних даних шляхом скорочення кількості змінних, зберігаючи при цьому максимум інформації. Серед цієї групи методів можна назвати а) *аналіз головних компонент (PCA), який проводиться для зменшення розмірності шляхом виділення нових, некорельованих змінних (головних компонент) та виходить з твердження, що перша компонента пояснює найбільшу частину дисперсії даних; б) факторний аналіз (FA) – проводиться для виділення латентних (прихованих) факторів, які впливають на змінні та передбачає припущення, що змінні корелюють через спільні фактори; в) мультимірне шкалювання (MDS) – проводиться для зменшення розмірності з урахуванням схожості між об'єктами;*

при цьому результат візуалізується у вигляді багатовимірної карти, що відображає відстані між об'єктами.

3. Методи виявлення зв'язків між змінними.

Ці методи дозволяють аналізувати залежності між змінними. Серед таких використовують: а) *канонічний кореляційний аналіз (ССА)* для вивчення зв'язків між двома наборами змінних; при цьому знаходять пари лінійних комбінацій змінних із максимальною кореляцією; б) *крос-кореляційний аналіз* для аналізу залежностей між змінними у часі; використовується для часових рядів; в) *моделювання структурних рівнянь (SEM)* для виявлення та перевірки причинно-наслідкових зв'язків між змінними; при цьому поєднуються елементи факторного й регресійного аналізу.

4. Методи для змішаних типів даних.

Серед таких методів виділяють: а) *кореспондентний аналіз* для вивчення взаємозв'язків між категоріальними змінними; при цьому результати представляються у вигляді графіка (двовимірного простору); б) *латентно-класовий аналіз* для ідентифікації груп у даних, які мають приховані класи; такий метод підходить для аналізу якісних даних.

Розглянемо більш детально деякі з найбільш поширених методів БСА. Кластерний аналіз – це метод багатомірного статистичного аналізу, який використовується для групування об'єктів (даних, спостережень) у групи або кластери так, щоб об'єкти всередині одного кластеру були схожими між собою, а об'єкти з різних кластерів – максимально відрізнялися. Кластерний аналіз належить до некерованого навчання, оскільки для його проведення не потрібно заздалегідь знати кількість груп чи їхню структуру. Цей метод широко застосовується у різних галузях науки, включаючи історичні дослідження, для аналізу великих обсягів даних і виявлення закономірностей.

Основні завдання кластерного аналізу полягають у: а) виділенні структур у наборі даних, які можуть бути неочевидними; б) розподілі об'єктів (наприклад, людей, документів, географічних об'єктів) на сегменти за спільними характеристиками; в) спрощенні

складної багатовимірної структури даних для кращої візуалізації та інтерпретації.

Існують 2 основні типи кластеризації – *ієрархічна кластеризація*, при якій створюється дерево (дендрограма), де об'єкти поступово об'єднуються або розділяються на групи, або *неієрархічна кластеризація*, при якій використовується фіксована кількість кластерів (наприклад, метод k-середніх).

При кластеризації визначення схожості між об'єктами ґрунтується на обраних метриках, таких як евклідова відстань (для кількісних даних), косинусна подібність (для текстових чи векторних даних), манхеттенська відстань (для даних із великим числом змінних).

Існує багато підходів до проведення кластеризації, серед яких найпопулярніші є такими: метод k-середніх (k-means), метод середніх зв'язків (average linkage), метод максимальної або мінімальної відстані.

У процесі кластеризації дані часто потрібно стандартизувати, оскільки змінні з великими значеннями можуть домінувати у визначенні схожості.

Вибір метрики відстані, способу визначення кластерів та початкових параметрів може суттєво вплинути на результати.

Результати кластерного аналізу зазвичай представляються у вигляді дендрограм, теплових карт або багатовимірних проєкцій.

Кластерний аналіз широко використовується в історичних науках, наприклад: а) аналіз демографічних даних – групування регіонів за рівнем народжуваності чи смертності у певний період; б) класифікація письмових джерел – виявлення схожих текстів або документів за змістом чи стилем; в) дослідження соціально-економічної історії – поділ країн чи регіонів на групи за економічними, соціальними чи культурними характеристиками.

Отже, кластерний аналіз є потужним інструментом для групування об'єктів за подібністю, виявлення закономірностей та спрощення складних наборів даних. Його застосування дозволяє дослідникам знаходити приховані зв'язки та структури, що

особливо є важливим в історичних і міждисциплінарних дослідженнях.

Дискримінантний аналіз (Discriminant Analysis) – це метод багатовимірного статистичного аналізу, який використовується для класифікації об'єктів у заздалегідь визначені групи. Його основна мета – знайти математичну функцію (дискримінантну функцію), яка дозволяє максимально чітко розділити групи на основі їх характеристик. Цей метод належить до підходів керованого навчання, оскільки вимагає знання про приналежність об'єктів до груп на етапі навчання моделі.

Основні завдання дискримінантного аналізу полягають у:
а) класифікації – визначенні групи, до якої належить новий об'єкт, на основі його характеристик; б) описі групових відмінностей – виявлення характеристик, які найбільше відрізняють групи одна від одної; в) прогнозуванні – побудові моделі для прогнозування належності об'єктів до груп; г) перевірці гіпотез – вивчення того, чи існують значущі відмінності між групами.

Розрізняють *лінійний дискримінантний аналіз (LDA)*, у якому передбачається, що групи розділяються лінійно, а змінні мають однакові дисперсії, та *квадратичний дискримінантний аналіз (QDA)*, при проведенні якого допускається нелінійне розділення та різні дисперсії для груп. *Дискримінантна функція* – це математичне рівняння, що має вигляд:

$$D_i = b_1x_1 + b_2x_2 + \dots + b_kx_k + c,$$

де D_i – значення дискримінантної функції для об'єкта i , x_1, x_2, \dots, x_k – значення змінних, b_1, b_2, \dots, b_k – коефіцієнти, c – константа. Об'єкти класифікуються в ту групу, для якої D_i є максимальним. Метод дозволяє аналізувати кілька змінних, але кількість дискримінантних функцій не може перевищувати $(m - 1)$, де m – кількість груп. При застосуванні дискримінантного аналізу слід дотримуватися вимог до даних: змінні мають розподілятися нормально, дисперсії змінних у групах повинні бути однаковими (для LDA), спостереження мають бути незалежними. Загалом,

дискримінантний аналіз схожий на багатовимірну регресію, але його мета – класифікація, а не оцінка залежностей.

Дискримінантний аналіз має певні переваги: це простота інтерпретації дискримінантних функцій, можливість використання для класифікації нових спостережень, ефективність за умов дотримання припущень (нормальності та однаковості дисперсій) та можливість одночасно аналізувати кілька змінних.

Втім, дискримінантний аналіз має деякі недоліки, а саме чутливість до припущень – до порушення нормальності або однаковості дисперсій може спотворити результати; погана адаптація до нелінійних взаємозв'язків – LDA не працює ефективно, якщо групи не розділяються лінійно; вплив мультиколінеарності – висока кореляція між змінними може призводити до нестабільності результатів.

Проведення дискримінантного аналізу складається з таких етапів:

1. Попередня підготовка даних, яка включає вибір змінних, що будуть використані у моделі, перевірку припущень (нормальності розподілу, однаковості дисперсій).

2. Побудова дискримінантної функції, яке проводиться шляхом оцінювання коефіцієнтів функції на основі навчальної вибірки.

3. Класифікація об'єктів, яка включає визначення належності об'єктів до груп на основі значень дискримінантної функції.

4. Оцінка точності класифікації, яка проводиться шляхом перевірки результатів за допомогою тестової вибірки або крос-валідації.

Дискримінантний аналіз застосовується в історичних дослідженнях при аналізі письмових джерел у процесі класифікації текстів за хронологічними або стилістичними характеристиками, при дослідженні демографічних даних при розподілі населення за соціальними чи економічними категоріями, в ході організації порівняльного аналізу історичних подій при виділенні факторів, що найбільше впливають на результат певних подій.

Отже, дискримінантний аналіз – це ефективний метод багатомірної статистики для класифікації об'єктів і вивчення групових відмінностей. Його застосування дозволяє не лише розподілити об'єкти за категоріями, а й виявити ключові характеристики, що визначають відмінності між групами. Однак цей метод потребує ретельної перевірки припущень і адекватності вибору змінних, щоб забезпечити точність і достовірність результатів.

Регресійні методи для багатомірних даних використовуються для аналізу залежності між однією або кількома залежними змінними (результатами) та одним або кількома предикторами (факторами). Ці методи дозволяють дослідникам моделювати складні взаємозв'язки в даних і робити прогнози. У багатомірних дослідженнях вони забезпечують гнучкість у роботі з різними типами змінних.

Розрізняють такі типи регресійних методів для багатомірних даних:

1. *Лінійна регресія*, що проводиться з метою моделювання лінійної залежності між однією залежною змінною та кількома незалежними змінними. В історичних дослідженнях, наприклад, використовується для аналізу впливу економічних чи соціальних факторів на демографічні тенденції.

2. *Багатовимірна лінійна регресія* (Multivariate Linear Regression), що проводиться для аналізу одночасної залежності кількох залежних змінних від кількох незалежних. Такий метод може використовуватися, наприклад, для дослідження зв'язку між соціально-економічними показниками (наприклад, доходом, рівнем освіти) та демографічними характеристиками (народжуваністю, смертністю).

3. *Логістична регресія* для моделювання залежності між якісною залежною змінною (двійковою чи багатокласовою) та кількома незалежними змінними. Такий метод підходить для класифікації об'єктів (наприклад, «належить» або «не належить»). Результати при цьому інтерпретуються як ймовірності. Може застосовуватися в

історичних дослідженнях для прогнозування ймовірності події на основі певних факторів (наприклад, ймовірності повстання залежно від соціально-економічних умов).

4. *Регресія часткових найменших квадратів (PLS Regression)* для аналізу залежності між матрицею залежних змінних і матрицею незалежних змінних, навіть за наявності сильно корельованих предикторів. Цей метод використовується при мультиколінеарності між змінними і поєднує зменшення розмірності та регресійний аналіз. Може застосовуватися при аналізі взаємозв'язку між різними економічними та соціальними показниками, що є сильно корельованими.

5. *Регресія для категоріальних даних* може використовуватися у вигляді поліноміальної логістичної регресії для багатокласових залежних змінних, у вигляді порядкової регресії для залежних змінних, що мають порядок, але не рівновіддаленість між рівнями (наприклад, рівень освіти: низький, середній, високий). Може застосовуватися при проведенні аналізу соціальних статусів чи рівнів освіти у різних історичних періодах.

Застосування регресійних методів містить такі етапи:

1. Формулювання гіпотези – визначення залежної змінної та предикторів.

2. Попередня обробка даних – перевірка даних на наявність пропусків, нормальності розподілу, мультиколінеарності.

3. Побудова моделі – оцінювання коефіцієнтів регресії за допомогою статистичних пакетів або програм.

4. Оцінка адекватності моделі – використання таких показників, як R^2 , AIC, BIC, тестів значущості.

5. Інтерпретація результатів – аналіз впливу кожного предиктора та прогнозування.

6. Перевірка моделі – використання тестової вибірки або крос-валідації для оцінки точності.

Регресійні методи активно використовуються для аналізу демографічних даних, а саме для моделювання залежності між економічними факторами та змінами народжуваності/смертності в

історичні періоди, для проведення соціально-економічних досліджень – для вивчення впливу рівня податків чи реформ на добробут населення, в історичній географії при оцінці впливу природних умов на розвиток міст чи регіонів.

Отже, регресійні методи для багатомірних даних є універсальним інструментом для вивчення складних взаємозв'язків. У контексті історичних досліджень вони дозволяють аналізувати вплив численних факторів на події чи явища, будувати прогностичні моделі та робити глибокі висновки про закономірності розвитку суспільств.

Аналіз головних компонент (Principal Component Analysis, PCA) – це метод зменшення розмірності даних, який дозволяє перетворити великий набір взаємозалежних змінних у менший набір нових незалежних змінних (головних компонент), зберігаючи при цьому максимально можливу частку інформації. PCA широко використовується для візуалізації багатовимірних даних, усунення кореляцій між змінними, зменшення обчислювальної складності та виявлення прихованих структур.

Основні цілі використання PCA є такими: а) зменшення розмірності даних шляхом спрощення даних зі збереженням основної частини їх варіативності; б) виявлення структур у даних через розкриття нових ознак, що описують приховані закономірності; в) усунення мультиколінеарності шляхом заміни корельованих змінних на незалежні компоненти; г) підготовка даних до інших аналізів через спрощення задач класифікації, кластеризації або регресійного аналізу.

Принципи роботи PCA полягають у: а) центруванні даних шляхом віднімання середнього значення кожної змінної, щоб дані мали середнє 0; б) обчисленні коваріаційної матриці, яка показує, як змінні співвідносяться одна з одною; в) знаходженні власних векторів і власних значень – власні вектори коваріаційної матриці визначають напрямки головних компонент, а власні значення – їх важливість (частку варіативності, яку пояснює кожна компонента); г) проектуванні даних на простір головних компонент.

Результатами застосування PCA є нові осі в просторі даних, на які проєктуються початкові точки, причому кількість компонент завжди менша або дорівнює кількості початкових змінних. Крім того, відсоток варіативності показує, яку частку загальної варіативності даних пояснює кожна компонента. Дані перетворюються у простір з меншою розмірністю, зберігаючи основну інформацію.

При застосуванні PCA для багатомірних даних стає можливою візуалізація багатомірних даних – зменшення розмірності до 2D або 3D для створення графічного представлення даних (наприклад, у дослідженні історичних джерел можна порівнювати характеристики текстів різних періодів у спрощеному просторі). Крім того, усунення мультиколінеарності замінює сильно корельовані змінні на незалежні компоненти (наприклад, в економічних дослідженнях, де показники (ВВП, дохід, споживання) можуть бути взаємопов'язані). Відбувається зменшення обчислювальної складності – скорочуються розмірності даних з великою кількістю змінних перед виконанням інших аналізів, таких як кластеризація чи регресія. Виявляються нові ознаки – створюються нові змінні, які є комбінацією початкових, але краще пояснюють закономірності в даних (наприклад, аналіз демографічних даних для створення комплексних індексів). В історичних дослідженнях проводиться аналіз змін у соціальних чи економічних показниках різних періодів, досліджуються текстові джерела для виділення тематичних напрямів.

Застосування методів PCA проводиться за такими етапами:

1. Попередня підготовка даних – центрування та нормалізація змінних.

2. Розрахунок головних компонент – використання програмного забезпечення (наприклад, Python, R, SPSS) для обчислення власних векторів і значень.

3. Вибір кількості компонент – використання графіка «коліна» (scree plot) або аналіз відсотка поясненої варіативності.

4. Проекція даних – перетворення початкових даних у простір головних компонент.

5. Інтерпретація та аналіз – аналіз нових компонент, візуалізація даних, використання результатів для подальшого аналізу.

Наведемо схематично приклад застосування PCA в історичних дослідженнях. Нехай перед дослідниками ставиться задача щодо необхідності проведення аналізу економічних показників різних регіонів у ХІХ столітті.

Процес розв'язування задачі складається з таких дій:

1. Збір даних про доходи, виробництво, демографію, податки.

2. Центрування та нормалізація даних.

3. Виконання PCA для виділення 2–3 головних компонент.

4. Інтерпретація компонент – перша компонента може відображати загальний економічний розвиток; друга – вплив демографічних факторів.

5. Побудова графіка, який показує групування регіонів за характеристиками.

Отже, метод PCA – це потужний інструмент для аналізу багатомірних даних. У контексті історичних досліджень він дозволяє узагальнювати складні дані, знаходити приховані закономірності та візуалізувати результати. Метод ефективний для аналізу великих наборів даних і є ключовим кроком у підготовці до інших видів статистичного аналізу.

Факторний аналіз (Factor Analysis, FA) – це метод статистичного аналізу, який використовується для зведення великого набору змінних до меншої кількості прихованих (латентних) факторів. Він допомагає виявити приховані структури, які пояснюють кореляції між змінними, і визначити основні впливові фактори. Цей метод широко застосовується в соціальних, економічних, історичних і психологічних дослідженнях для виявлення прихованих закономірностей.

Мета проведення факторного аналізу полягає у: а) *виявленні прихованих факторів* через визначення основних характеристик, що

впливають на набір взаємозалежних змінних; б) скороченні розмірності шляхом зменшення кількості змінних, при якому зберігається основна частина їхньої інформації; в) усуненні мультиколінеарності через заміну корельованих змінних на незалежні латентні фактори; г) інтерпретації даних шляхом спрощення розуміння складних багатомірних даних.

На практиці застосовуються такі типи факторного аналізу:

1. *Експлораторний факторний аналіз (EFA)* – використовується для вивчення даних без попередньо сформульованих гіпотез щодо структури факторів.

2. *Конфірматорний факторний аналіз (CFA)* – використовується для перевірки гіпотез щодо факторної структури даних.

Проведення факторного аналізу має відповідати таким основним принципам: а) при виборі моделі факторного аналізу спостережувані змінні розглядаються як комбінація; б) обернення факторної матриці проводиться через ротацію факторів (ортогональна або косокутна), яка застосовується для спрощення інтерпретації факторів; в) для вибору кількості факторів використовується критерій власних значень (наприклад, правило Кайзера: фактори з власним значенням більше 1) або графік "коліна" (scree plot).

Застосування факторного аналізу проводиться за такими етапами:

1. *Збір і підготовка даних*, які включають перевірку на нормальність розподілу, мультиколінеарність, наявність пропусків.

2. *Обчислення кореляційної матриці* шляхом визначення кореляцій між змінними для виявлення взаємозалежностей.

3. *Вибір кількості факторів* через визначення оптимальної кількості факторів за допомогою власних значень або інших критеріїв.

4. *Обчислення факторних навантажень* шляхом використання алгоритмів (метод головних компонент, максимальна правдоподібність тощо).

5. *Ротація факторів* – ортогональна (наприклад, Varimax) або косокутна (наприклад, Promax) – для полегшення інтерпретації.

6. *Інтерпретація факторів* через визначення змісту кожного фактора на основі факторних навантажень.

7. *Перевірка моделі* шляхом аналізу залишків та проведення перевірки відповідності моделі даним.

Можливості застосування факторного аналізу сягають різноманітних галузей досліджень.

В історичних дослідженнях факторний аналіз застосовується при організації аналізу текстів шляхом виділення основних тематичних напрямів у документах різних епох, у соціальних дослідженнях для визначення основних факторів, що впливали на рівень добробуту в різні історичні періоди, при виявленні спільних характеристик культурних явищ у різних регіонах.

У демографії та економіці факторний аналіз застосовується для визначення основних показників, що впливають на демографічні зміни (народжуваність, смертність), а також для вивчення економічних чинників у розвитку регіонів.

Факторний аналіз знаходить своє застосування у соціології для аналізу суспільної думки, виділення основних напрямів у ставленні до історичних подій.

У психології цей метод використовується для виявлення латентних характеристик особистості на основі тестів.

Наведемо приклад застосування факторного аналізу в історичних дослідженнях. Нехай перед дослідниками стоїть задача дослідити ключові фактори, що впливали на соціально-економічний розвиток регіонів у XIX столітті. Процес розв'язання задачі включає такі етапи:

1. Збір даних, а саме показників – рівня грамотності, середнього доходу, кількості шкіл, показників міграції.

2. Проведення факторного аналізу, в результаті якого виявлено три основні фактори – освітній розвиток, показник якого має високу кореляцію з кількістю шкіл і рівнем грамотності;

економічний розвиток – кореляція з середнім доходом; міграційна активність – кореляція з числом міграцій.

3. Формулювання висновків розв'язання задачі, в яких стверджується, що ідентифіковані фактори дозволяють простежити роль освіти, економіки та міграції у формуванні регіональної специфіки.

Отже, факторний аналіз – це ефективний інструмент для виявлення прихованих структур у багатовимірних даних. У контексті історичних досліджень він дає змогу глибше зрозуміти складні соціальні, економічні чи культурні процеси минулого, виділивши основні фактори, що формували їх розвиток.

Метод мультимірного шкалювання (MDS) – це статистичний метод, який використовується для візуалізації відношень між об'єктами на основі подібності чи відстаней між ними в багатовимірному просторі. MDS дозволяє відобразити ці об'єкти в просторі меншої розмірності (зазвичай дво- або тривимірному), зберігаючи їхні відношення максимально точно.

Мета застосування методу MDS полягає у: а) перетворенні матриці відстаней або подібностей між об'єктами у візуально зрозумілу форму; б) ідентифікації латентних структур, які пояснюють відношення між об'єктами; в) спрощення аналізу складних багатомірних даних.

Основними принципами методу MDS є: а) MDS починається з квадратної симетричної матриці, яка містить міри подібності (кореляції, коефіцієнти Жаккара тощо) або відстані (евклідова, манхеттенська, косинусна відстань) між об'єктами; б) методом передбачається визначення простору меншої розмірності (зазвичай 2D або 3D), у якому об'єкти будуть відображені; в) метод MDS намагається розташувати об'єкти так, щоб відстані між ними у спрощеному просторі максимально відповідали відстаням у вихідному просторі; г) стрес-функція використовується для оцінки якості відображення – вона визначає, наскільки добре розташування у просторі меншої розмірності відповідає вихідним відстаням.

Розрізняють такі типи мультимірного шкалювання:

1. *Метричне MDS*, яке підходить для даних, де відстані мають метричні властивості, і зберігає пропорції між відстанями.

2. *Неметричне MDS*, яке використовується для рангових даних або подібностей, зберігаючи при цьому порядок об'єктів, а не точні відстані.

Етапи застосування MDS у загальному випадку визначаються так:

1. *Збір даних* – формування матриці подібностей або відстаней. При цьому дані можуть бути отримані з історичних, соціальних або культурних досліджень.

2. *Підготовка матриці* – перевірка матриці на симетричність і відсутність пропусків. Проведення за необхідності нормалізації даних.

3. *Вибір типу MDS* – метричне або неметричне шкалювання залежно від характеру даних.

4. *Виконання аналізу* – використання спеціалізованих програм (SPSS, R, Python, MATLAB) для обчислення координат об'єктів у просторі меншої розмірності.

5. *Оцінка якості моделі* – аналіз стрес-функції. Значення менше 0,1 вважається добрим, а значення понад 0,2 – незадовільним.

6. *Візуалізація результатів* – побудова графіків або діаграм, що показують розташування об'єктів у зменшеному просторі.

7. *Інтерпретація* – аналіз відношень між об'єктами на основі їхнього розташування, визначення кластерів, відстаней і можливих закономірностей.

Застосування MDS у історичних дослідженнях може включати такі варіанти: а) аналіз культурних взаємозв'язків шляхом вивчення подібності між різними культурами чи регіонами за показниками розвитку, традицій чи мистецтва; б) визначення розташування історичних подій через побудову візуальних моделей взаємозв'язків між подіями на основі їхніх характеристик (географія, хронологія, соціально-політичні фактори); в) дослідження мовних подібностей шляхом аналізу історичних змін мов та їхньої подібності для

відображення мовного родоводу; г) реконструкція соціальних структур шляхом вивчення взаємовідносин між соціальними групами, професіями чи класами в різні епохи; д) порівняння регіонального розвитку шляхом побудови карт, що демонструють соціально-економічні подібності чи відмінності між регіонами.

Наведемо схематичний приклад застосування MDS, який полягатиме у розв'язанні задачі вивчення подібності між країнами Європи у XIX столітті за показниками економічного розвитку.

Процес розв'язання задачі включає такі етапи:

1. Визначення даних величин ВВП, рівня урбанізації, чисельності робітничого класу.
2. Створення матриці відстаней - розрахунок евклідової відстань між країнами за обраними показниками.
3. Застосування методу MDS шляхом виконання неметричного шкалювання.
4. Отримання результату у вигляді карти, де країни з подібним рівнем розвитку розташовані близько одна до одної.
5. Формулювання висновку, що групи країн (наприклад, індустріально розвинені чи аграрні) виділяються в окремі кластери, що відображає загальні тенденції економічного розвитку.

Отже, метод мультимірного шкалювання є потужним інструментом для візуалізації багатовимірних даних і виявлення закономірностей у взаємовідносинах між об'єктами. У контексті історичних досліджень він дозволяє дослідникам краще зрозуміти складні соціальні, економічні чи культурні процеси, що відбувалися у минулому.

Загалом, застосування багатомірного статистичного аналізу в історії включає кілька послідовних етапів, що забезпечують якісний і коректний аналіз даних. Кожен з цих етапів є критично важливим для досягнення точних і релевантних результатів.

1. Постановка дослідницької задачі.

На цьому етапі визначається мета дослідження (які історичні процеси, події або явища необхідно проаналізувати), ключові питання (які залежності чи закономірності між історичними

змінними потрібно встановити), гіпотези (попередні припущення про характер взаємозв'язків у даних).

2. Збір і підготовка даних.

Цей етап включає збір даних (архівні джерела, статистичні звіти, демографічні дані, історичні описи), оцифрування або введення даних у зручний формат (електронні таблиці, бази даних), попередній аналіз (виявлення пропусків у даних, перевірка точності та надійності джерел), масштабування змінних (нормалізація або стандартизація для забезпечення порівнянності даних), категоризація даних (перетворення якісних змінних у кількісні (кодування), якщо це потрібно).

3. Вибір методу аналізу.

Залежно від характеру даних і мети дослідження обирається відповідний метод БСА – кластерний аналіз для групування подій, територій або об'єктів, дискримінантний аналіз для визначення відмінностей між групами, аналіз головних компонент (РСА) для зменшення розмірності даних, факторний аналіз для виявлення латентних змінних, регресійні методи для виявлення залежностей між змінними, мультивимірна шкалярна регресія для просторової візуалізації даних.

4. Проведення попереднього аналізу.

За допомогою описової статистики проводяться обчислення середніх значень, стандартних відхилень, перевірка розподілів. Організовується перевірка кореляцій для оцінки взаємозв'язків між змінними. Виявляється мультиколінеарність шляхом перевірки наявності корельованих змінних. Проводиться аналіз аномалій: для виявлення та обробка викидів.

5. Виконання основного аналізу.

Організація залежно від обраного методу БСА формування матриці даних для підготовки даних у вигляді, придатному для обчислень, розрахунків результатів за допомогою програмних засобів (SPSS, Python, R, Stata тощо), перевірка адекватності моделі через формування оцінки відповідності моделі даним.

6. Інтерпретація результатів.

Проведення опису факторів чи груп для пояснення змісту кластерів, головних компонент або факторів. Створення графіків, діаграм, дендрограм, карт для візуалізації даних. Співставлення отриманих результатів із контекстом історичного періоду.

7. Перевірка достовірності результатів.

Проводиться перевірка відповідності моделі даним (аналіз залишків), перевірка стійкості результатів на іншій вибірці (крос-валідація), оцінка узгодженості з відомими історичними фактами чи іншими дослідженнями (порівняння з іншими дослідженнями).

8. Узагальнення результатів і висновки.

Формування підсумкових висновків щодо того, які історичні закономірності були виявлені. Формування рекомендацій для подальших досліджень – які аспекти варто дослідити глибше.

9. Представлення результатів.

Підготовка звітів, статей, презентацій – оформлення даних у науково прийнятному вигляді. Використання результатів в освітньому чи науковому процесі з демонстрацією висновків і методів дослідження.

Таким чином, проведення багатомірного статистичного аналізу в історичних дослідженнях є багатоступеневим процесом, який включає підготовку, аналіз і інтерпретацію даних. Застосування таких методів дозволяє виявити приховані закономірності, які складно було б побачити традиційними способами аналізу.

Багатомірний статистичний аналіз активно використовується в історичних дослідженнях для виявлення закономірностей у великих масивах даних. Наведемо прикладі успішних кейсів, які ілюструють ефективність цих методів.

1. Дослідження на тему «Аналіз соціальної стратифікації в середньовічній Європі» [23]. Застосовувався при проведенні дослідження кластерний аналіз. В якості даних був проведений зріз соціально-економічних характеристик різних груп населення в середньовічних європейських містах (зайнятість, доходи, власність на землю).

В результаті проведення дослідження було виокремлено основні соціальні класи, включаючи дворянство, ремісників і селян; встановлено, що економічна мобільність була найвищою у великих міських центрах, що свідчило про ранні прояви міського капіталізму.

Таким чином, кластерний аналіз дозволив візуалізувати соціальні структури і зміни, що стали передвісниками формування сучасного класового суспільства.

2. Дослідження на тему «Географічна реконструкція торгових шляхів у добу Відродження»[22]. Про проведенні дослідження застосовувався аналіз головних компонент (РСА). В якості даних були взяті обсяги торгівлі товарами (тканинами, спеціями, металами) між європейськими містами за архівними документами.

В результаті проведення дослідження виділено два основні фактори, що впливали на формування торгових шляхів – близькість до основних річкових і морських маршрутів та політична стабільність регіону. Зменшення розмірності даних дозволило простежити зв'язок між економічним зростанням міст і їхнім розташуванням у торговельній мережі.

Таким чином, використання методу РСА дозволило виявити ключові рушії економічного розвитку, що сприяло розумінню формування мережі європейської торгівлі.

3. Дослідження на тему «Мовні подібності в слов'янських мовах»[23]. Застосовувався метод мультимірного шкалювання (MDS). В якості даних бралася частота вживання коренів слів і граматичних конструкцій у слов'янських мовах.

В результаті проведення дослідження реконструйовано «мовну карту» слов'янських мов, де мови, що мали спільне походження, виявилися ближчими одна до одної. Дослідження підтвердило, що українська та білоруська мови мають найближчі зв'язки з давньоруською мовою.

Отже, метод MDS став інструментом для кращого розуміння мовних і культурних взаємозв'язків.

4. Дослідження на тему «Аналіз причин революційних рухів у ХІХ столітті» [19]. В процесі аналізу застосовувався регресійний аналіз. Для проведення дослідження були зібрані політичні, економічні та демографічні показники країн, де відбулися революційні події (частка населення, зайнятого у промисловості, рівень грамотності, податкове навантаження).

В результаті проведення дослідження встановлено, що основними детермінантами революцій були високий рівень урбанізації та значні соціальні нерівності. Низький рівень участі у виборах корелював із більшою ймовірністю заворушень.

Таким чином, регресійний аналіз дозволив побудувати ретропрогнозну модель і перевірити ключові теорії про революції.

5. Дослідження на тему «Реконструкція демографічних змін у період пандемії «чорної смерті»[30]. Застосовувався при проведенні дослідження факторний аналіз. Для його проведення були зібрані дані про рівні смертності, відновлення населення, економічні показники в європейських регіонах ХІV століття.

В результаті проведення дослідження було виділено основні фактори, які впливали на швидкість демографічного відновлення: щільність населення, розвиток аграрного сектору та торгівлі. Було також встановлено, що регіони зі змішаною економікою (аграрно-торговельні) відновлювалися швидше, ніж суто аграрні.

Таким чином, факторний аналіз допоміг зрозуміти, чому одні регіони швидше оговталися від кризи, ніж інші.

6. Проєкт «Дослідження міграційних процесів у ХХ столітті»[31], реалізація якого проводилася з використанням кластерного аналізу в поєднанні з регресійними моделями. Для проведення дослідження була зібрана статистика міграції, рівень безробіття, економічна привабливість регіонів.

В результаті виконання проєкту виділено кластери регіонів за типами міграційних процесів: а) еміграційні (з країн із низьким ВВП); б) імміграційні (економічно розвинені країни); в) транзитні регіони. Регресійний аналіз показав, що ключовими рушіями міграції були економічні стимули та політична стабільність.

Як наслідок, аналіз міграційних потоків дав змогу прогнозувати подібні процеси у майбутньому.

7. Дослідження на тему «Карта релігійної подібності в ранньомодерній Європі»[12] проводилося з використанням методу мультимірного шкалювання (MDS). В якості даних були зібрані тексти релігійних документів, кількісні показники поширення конфесій.

В результаті проведення дослідження побудовано карту, яка відображає конфесійні зв'язки між регіонами. При цьому регіони з домінуванням протестантизму утворили один кластер, католицизму – інший, а православ'я – окрему групу.

Таким чином, дослідження показало динаміку релігійної сегрегації та її зв'язок із політичними та економічними факторами.

8. Дослідження на тему «Просторовий аналіз військових кампаній» [31] проводилося з використанням аналізу головних компонент у поєднанні з географічним моделюванням. В якості даних були зібрані географічні координати битв, чисельність військ, втрати.

В результаті проведення дослідження виявлено ключові фактори, що впливали на успішність військових кампаній: рельєф, доступність ресурсів, погодні умови. Карти кампаній стали наочними інструментами для розуміння стратегій ведення війни.

9. Проект на тему «З'їзди й конгреси природознавців і математиків у «віснику дослідної фізики та елементарної математики» (1886–1917 рр.): аналіз публікацій» [11], у якому представлені результати дослідження публікацій у «Віснику дослідної фізики та елементарної математики», що видавався в Києві та Одесі протягом 1886–1917 рр., про організацію, проведення та результати вітчизняних і закордонних з'їздів та конгресів математиків і природознавців. Журнал був неофіційним періодичним друкованим виданням математичного відділення Новоросійського товариства дослідників природи. Дослідження проведене з метою здійснення змістовного та кількісного аналізу текстів публікацій журналу, в яких висвітлюються матеріали таких зібрань науковців та педагогів. Авторами для змістовного аналізу предмету

дослідження використовувалися наукові методи, а в процесі кількісного аналізу – квантифікація тексту, збирання емпіричних даних, їхнє узагальнення та математико-статистична обробка.

У результаті проведеного дослідження матеріалів журналу про з'їзди та конгреси математиків і дослідників природи протягом всього періоду його видання автори дійшли таких висновків.

По-перше, матеріали журналу про з'їзди та конгреси за всі роки його видання становили трохи більше 4 % його загальної площі.

По-друге, всі вітчизняні з'їзди природодослідників та математиків, які збиралися протягом періоду видання журналу, були висвітлені на його сторінках (це майже 2/3 від всього обсягу інформації журналу про такі зібрання науковців та педагогів). При цьому способи подання інформації та її обсяг суттєво відрізнялися за різними зібраннями в залежності від наявності мотивів обох складів редакційної колегії. Мотивами розміщення інформації були: а) наявність в роботі зібрань педагогічної складової; б) наукова (або педагогічна) значимість доповідей та учасників зібрання; в) доступність якісних матеріалів про зібрання; г) наявність журнального місця. Міжнародні з'їзди та конгреси висвітлювалися вісником нерегулярно, мотиви для розміщення в журналі інформації про такі зібрання були аналогічними.

По-третє, починаючи з 1901 року редакція збірника впровадила до практики представлення матеріалів про наукові й педагогічні з'їзди та конгреси відпрацьовану протягом попередніх років існування журналу схему: а) оголошення про подію; б) публікація положення (статуту, програми) події; в) опис підготовки до події; г) огляду особливостей проведення події; д) презентація текстів важливих виступів. Така схема презентації зібрань була запроваджена вперше в науково-популярній та навчальній періодиці Російської імперії.

По-четверте, педагогічна складова була найбільш вагомим мотивом для обох складів редакції Вісника при прийнятті рішення щодо особливостей та обсягів висвітлення матеріалів про роботу вітчизняних та закордонних з'їздів і конгресів.

Таким чином, багатомірний аналіз доповнив історичні дослідження, надавши нове бачення суспільних процесів, військових стратегій, змін у науково-популярній літературі та мовному середовищі. Втім, застосування багатомірного статистичного аналізу в історичних дослідженнях часто супроводжується рядом труднощів, які пов'язані зі специфікою історичних даних. Основні проблеми можуть бути такими.

1. Історичні дані часто є неповними та фрагментарними через втрату джерел, неточності у записах або їх вибірковість. Це ускладнює побудову надійних моделей, адже пропуски в даних можуть спотворювати результати аналізу. Для уникнення такого роду проблем застосовуються методи реконструкції даних (наприклад, використання середніх значень, прогностичних моделей або експертних оцінок).

2. Дані, отримані з різних історичних джерел, можуть бути непорівнюваними та мати розбіжності у форматах, одиницях вимірювання або інтерпретаціях. Як наслідок, такі дані складно поєднувати та аналізувати в межах одного дослідження. Рішенням цієї проблеми може бути застосування нормалізації даних або переведення їх до єдиного формату.

3. Дані, що збереглися, могли бути упередженими та суб'єктивними, адже їх створювали люди з власними інтересами чи поглядами. Це може значно спотворювати аналіз і призводити до неправильних висновків. В даній ситуації допомогти може перевірка даних з різних джерел, критичний підхід до оцінки достовірності та незалежності даних.

4. На точність даних можуть впливати часові розриви та змінність даних. Це пов'язано з тим, що історичні події мають тривалість у часі, і характеристики явищ можуть змінюватися. Як наслідок, це ускладнює визначення сталих закономірностей та трендів. Рішенням для організації якісного дослідження таких проєктів може полягати в розподілі аналізу на часові періоди або використання методів аналізу часових рядів.

5. Історичні дані часто мають багато змінних, що ускладнює їх обробку. Таким чином, висока вимірність та складність багатомірних даних може призводити до перевантаження моделей даними (ефект «curse of dimensionality»). В таких проєктах рішенням може стати використання методів зменшення вимірності, таких як аналіз головних компонент (PCA) або факторний аналіз.

6. Результати багатомірного аналізу часто мають математичний характер, що потребує додаткової чіткої інтерпретації в історичному контексті. При цьому ймовірною є можливість отримання помилкових або надмірно узагальнених висновків. Рішення цієї проблеми полягає у налагодженні співпраці з професійними істориками, які можуть надати якісний аналіз отриманих числових результатів.

7. Застосування багатомірного статистичного аналізу в історичних дослідженнях серйозно залежать від технічного та програмного забезпечення – для проведення багатомірного аналізу потрібні спеціалізовані програми (SPSS, R, Python тощо) та потужне обладнання. Недостатність технічних ресурсів може обмежити можливості дослідження. Тому в таких проєктах слід максимально організувати використання хмарних сервісів, адаптованого програмного забезпечення та різні форми навчання персоналу.

8. Багатомірний аналіз може виявляти кореляції, які насправді є випадковими і не відображають причинно-наслідкових зв'язків. Це може призводити до неправильної інтерпретації даних. Для запобігання таких ситуацій слід організовувати перевірку результатів на різних вибірках, а також використовувати додаткових методів для підтвердження гіпотез.

9. Нарешті, очевидним є факт, що сучасні ідеї суттєво впливають на інтерпретацію результатів досліджень, так як всі історичні дані аналізуються крізь призму сучасних уявлень. Це може призводити до викривлення історичного контексту і, як наслідок, втрати автентичності в результатах. Застосування міждисциплінарного підходу та врахування історичних реалій

досліджуваної епохи може допомогти дослідниками правильно реагувати на такого роду загрози.

Таким чином, попри ці труднощі, багатовимірний статистичний аналіз залишається потужним інструментом, що дозволяє розкрити складні історичні закономірності. Успішне застосування потребує ретельного планування, адаптації методик та критичного підходу до інтерпретації результатів. Методи багатомірного статистичного аналізу мають низку сильних сторін, які роблять їх ефективними в історичних дослідженнях. Основні переваги таких методів можна назвати так.

1. Для методів БСА характерними є *об'єктивність та кількісний підхід*, так як вони дозволяють проводити аналіз великих масивів даних, виключаючи суб'єктивні упередження дослідника. Це важливо при вивченні складних історичних процесів, які містять безліч факторів.

2. Завдяки методам БСА, як-от аналіз головних компонент (РСА) або факторний аналіз, можна *виявляти приховані закономірності*, взаємозв'язки та структури в даних. Наприклад, такі підходи дозволяють виявити фактори, що впливали на економічний розвиток різних регіонів у певний період.

3. Методи БСА дають змогу *проводити порівняльний аналіз міжрегіональних та міжчасових особливостей явищ та процесів* на основі узгоджених кількісних критеріїв. Це важливо для порівняння культурних, економічних або соціальних явищ між різними епохами чи регіонами.

4. Ці методи допомагають *структурувати та узагальнювати великі обсяги даних*, які є складними та багатовимірними. Наприклад, використання кластерного аналізу для групування історичних подій за спільними характеристиками.

5. Застосування методів регресійного аналізу або мультимірного шкалювання дозволяє *моделювати історичні процеси*, їхні причинно-наслідкові зв'язки, прогнозувати тенденції та реконструювати історичні події.

6. Методи БСА ефективно інтегруються з сучасними інструментами, такими як геоінформаційні системи (ГІС), що дозволяє поєднувати просторові, демографічні та соціальні аспекти історичних даних.

7. Методи БСА здатні обробляти великі масиви даних, зокрема оцифровані архіви, переписи, економічні звіти тощо, що значно розширює можливості історичних досліджень.

8. В арсеналі БСА є методи для аналізу залежностей (дискримінантний аналіз), виявлення кластерів (кластерний аналіз), зменшення вимірності (РСА) та інші, що дозволяє адаптувати підхід до конкретного завдання.

9. Методи БСА (наприклад, мультимірне шкалювання (MDS)) дозволяють візуалізувати багатовимірні дані у вигляді графіків, спрощуючи інтерпретацію складних структур і відношень.

10. Методи БСА мають міждисциплінарний потенціал, так як дозволяють інтегрувати історичні дані з результатами інших наук (соціології, економіки, демографії), розширюючи методологічні рамки дослідження.

Загалом, методи БСА є незамінним інструментом для істориків, які прагнуть поєднати традиційні методи аналізу з сучасними технологічними підходами. Подальший розвиток методології застосування багатомірного статистичного аналізу в історичних дослідженнях відкриває значні перспективи завдяки поєднанню традиційного історичного аналізу з сучасними науковими технологіями за такими напрямками.

1. *Розширення джерельної бази:* а) цифровізація історичних джерел – зростання доступу до оцифрованих архівів, баз даних та документів дозволяє створювати великі вибірки для аналізу, що сприятиме використанню більш складних багатомірних моделей; б) інтеграція даних різного типу – поєднання текстових, просторових, графічних і числових даних для комплексного аналізу історичних явищ.

2. *Поглиблення міждисциплінарного підходу:* а) поєднання з іншими науками – інтеграція методів із соціології, економіки,

археології та екології дозволяє отримувати нові інсайти, наприклад, про соціальні структури, міграційні процеси або вплив екологічних факторів на історичні події; б) розвиток концепцій «Digital Humanities» – використання сучасного програмного забезпечення (R, Python, GIS) сприятиме кращій обробці даних.

3. *Розвиток алгоритмів і моделей*: а) поліпшення алгоритмів кластерного аналізу (наприклад, адаптація до роботи з нерівномірними даними та розробка нових метрик відстаней); б) застосування методів машинного навчання – використання алгоритмів штучного інтелекту для пошуку прихованих закономірностей і прогнозування історичних тенденцій; в) впровадження нелінійних моделей, що може покращити аналіз складних соціальних та економічних процесів у довгостроковій перспективі.

4. *Візуалізація та інтерпретація результатів*: а) використання інтерактивних візуалізацій – нові технології дозволяють представляти результати аналізу у вигляді інтерактивних карт, графіків і динамічних моделей, що полегшує їх інтерпретацію; б) гуманітарно-орієнтована візуалізація – створення інструментів, які враховують специфіку історичних даних, таких як часові шкали або історичні мапи.

5. *Застосування великих даних (Big Data)*: а) аналіз великих масивів історичної інформації, таких як переписи населення, економічні звіти чи епістолярні джерела; використання при цьому БСА допомагає виявляти тренди на макрорівні.

6. *Вивчення впливу невизначеності в даних*: а) аналіз нечітких даних – розробка методик для врахування неповноти чи неточності в історичних даних (наприклад, застосування нечіткої логіки або баєсових методів); б) розрахунок надійності результатів – використання статистичних підходів для оцінки стабільності висновків на основі частково реконструйованих даних.

7. *Розвиток спеціалізованих програмних платформ*: створення програмних продуктів, адаптованих до потреб істориків

(наприклад, для аналізу текстових даних або візуалізації багатовимірних моделей у контексті часових і просторових змін).

8. *Етичний аналіз історичних даних*: застосування багатовимірного аналізу для етичного вивчення делікатних історичних тем, таких як геноциди чи рабство, із врахуванням впливу культурних упереджень на створення та збереження джерел.

9. *Популяризація результатів*: використання інтерактивних засобів для популяризації історичних досліджень серед широкої аудиторії, зокрема в освітніх програмах чи музеях.

Таким чином, методи багатомірного статистичного аналізу відкривають нові можливості для дослідження складних історичних явищ, які часто включають велику кількість взаємопов'язаних змінних. Їх використання забезпечує комплексний підхід до аналізу даних, дозволяючи виявляти приховані закономірності, моделювати динаміку історичних процесів та робити міжчасові та міжрегіональні порівняння.

Ключові методи, такі як аналіз головних компонент, факторний, кластерний аналіз, дискримінантний аналіз та мультимірне шкалювання, дозволяють історикам працювати з великими обсягами даних, інтегруючи кількісний і якісний підходи. Завдяки їх використанню можна структурувати дані, виявляти ключові фактори впливу, а також прогнозувати тенденції, що сприяє глибшому розумінню історичних процесів.

Сучасні наукові дослідження демонструють ефективність БСА у вивченні соціальних, економічних, демографічних і політичних явищ. Наприклад, успішне застосування цих методів у дослідженнях динаміки урбанізації, соціальної стратифікації або еволюції політичних структур свідчить про їхню адаптивність до різних типів історичних даних.

Попри це, застосування БСА в історичних дослідженнях супроводжується певними викликами. Серед основних – якість і повнота вихідних даних, методологічні труднощі в інтерпретації результатів та необхідність міждисциплінарного підходу. Однак

впровадження новітніх технологій, таких як машинне навчання та геоінформаційні системи, поступово долає ці бар'єри.

У підсумку, використання методів багатомірного статистичного аналізу є важливим етапом модернізації історичних досліджень, що сприяє їхній науковій об'єктивності, міждисциплінарності та інтеграції з іншими галузями знань. Це дозволяє історикам відкривати нові горизонти у вивченні минулого, будуючи науково обґрунтовані висновки на основі кількісного аналізу.

8.2. Застосування інструментів EXCEL для багатомірного статистичного аналізу

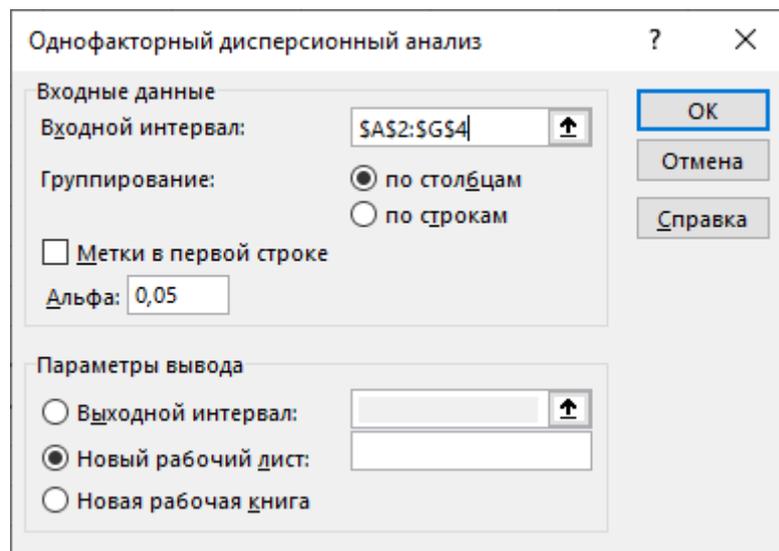
Розглянемо задачу визначення впливу форми економічного укладу (1 – аграрний, 2 – змішаний, 3 – промисловий) європейських країн XIV століття на відновлення населення протягом 5-ти років після пандемії. Проведемо для цього однофакторний дисперсійний аналіз. Цей засіб виконує простий аналіз дисперсії даних для кількох зразків. Аналіз дає змогу перевірити гіпотезу, що кожна вибірка взята з одного базового розподілу ймовірності проти альтернативної гіпотези про те, що базові розподіли ймовірності не однакові для всіх вибірок.

Побудуємо в пакеті MSEXCEL відповідну таблицю даних за країнами.

	A	B	C	D	E	F	G	H
1	фактори	країна 1	країна 2	країна 3	країна 4	країна 5	країна 6	
2	1	14	20	16	22	32	37	
3	2	16	24	19	28	34	40	
4	3	12	22	15	26	32	36	
5								

Відкриваємо діалогове вікно аналітичного інструмента «Аналіз даних». У списку вибираємо «Однофакторний дисперсійний аналіз» і натискаємо ОК. У полі «Вхідний інтервал» вводимо посилання на

діапазон комірок, що містяться у всіх стовпцях таблиці. «Групування» призначаємо за стовпцями. Альфа (рівень значущості) дорівнює 0,05. «Параметри виведення» – новий робочий лист. Якщо потрібно вказати вихідний діапазон на наявному аркуші, то перемикач ставимо в положення «Вихідний інтервал» і посилаємося на ліву верхню комірку діапазону для даних, що виводяться. Розміри визначаються автоматично. Після встановлення параметрів натисніть кнопку (ОК). Екран налаштувань має виглядати так:



Результати факторного дисперсійного аналізу будуть представлені на новому аркуші та складаються з двох таблиць. У першій таблиці для кожного рядка та кожного стовпця вихідної таблиці наведено числові параметри: кількість чисел, сума, середня та дисперсія.

	A	B	C	D	E	F	G
1	Однофакторный дисперсионный анализ						
2							
3	ИТОГИ						
4	<i>Группы</i>	<i>Счет</i>	<i>Сумма</i>	<i>Среднее</i>	<i>Дисперсия</i>		
5	Столбец 1	3	6	2	1		
6	Столбец 2	3	42	14	4		
7	Столбец 3	3	66	22	4		
8	Столбец 4	3	50	16,66666667	4,333333		
9	Столбец 5	3	76	25,33333333	9,333333		
10	Столбец 6	3	98	32,66666667	1,333333		
11	Столбец 7	3	113	37,66666667	4,333333		
12							
13							
14	Дисперсионный анализ						
15	<i>Источник вариации</i>	<i>SS</i>	<i>df</i>	<i>MS</i>	<i>F</i>	<i>P-Значение</i>	<i>F критическое</i>
16	Между группами	2582,571	6	430,4285714	106,3412	7,2908E-11	2,847725996
17	Внутри групп	56,66667	14	4,047619048			
18							
19	Итого	2639,238	20				
20							

Друга частина – це результати дисперсійного аналізу. У таблиці результатів Excel використовує такі позначення:

SS – сума квадратів,

df – ступені свободи,

MS – середній квадрат (дисперсія),

F – значення F-статистики Фішера, що спостерігається,

p-значення – значимість критерію Фішера,

F-критичне – критичне значення F-статистики при $p=0,05$.

Проаналізуємо отримані результати. Відповідно до другої таблиці маємо $F=106,3412$ і $F_{кр.}=2,848$, тобто, $F > F_{кр.}$. Отже вплив чинника на результативну ознаку вважатиметься достовірним, що й підтверджується величиною значимості ($p=7,3 \cdot 10^{-11}$), що менше 0,05.

Отже, конкуруюча гіпотеза про достовірність впливу фактора може бути прийнята з ймовірністю 0.95. При цьому сила впливу фактора форми економічного укладу європейських країн XIV століття на відновлення населення протягом 5-ти років після пандемії оцінюється так:

$$K_A = \frac{2582.571}{2639.238} * 100\% = 97.8\%.$$

Сформулюємо іншу задачу. Необхідно методом дисперсійного аналізу оцінити вплив форми економічного укладу (1 – аграрний, 2 – змішаний, 3 – промисловий) європейських країн XIV століття та географічного розміщення цих країн на смертність їхніх жителів під час пандемії. Відповідно до таблиці фактор А має 3 рівні (1–3), а фактор В – 6 рівнів (країни 1–6). У кожній групі є лише по одному спостереженню. Будемо використовувати двофакторний дисперсійний аналіз без повторень. Вихідні дані представлені у таблиці.

	A	B	C	D	E	F	G
1	смертність (%)						
2	фактори	країна 1	країна 2	країна 3	країна 4	країна 5	країна 6
3	1	10	11	12	11	14	10
4	2	11	12	15	16	16	9
5	3	14	13	16	15	19	11

Виконуємо команду *Аналіз даних* із меню *Дані*. Відповідно до умов завдання вибираємо у діалоговому вікні метод «Двофакторний дисперсійний аналіз без повторень» і натискаємо кнопку (ОК). У вікні «Двофакторний дисперсійний аналіз без повторень» встановимо для вхідних даних такі параметри: вхідний інтервал дорівнює $B3:G5$, вхідний діапазон не містить мітки, альфа (рівень значимості) дорівнює 0,05. Для параметрів виводу встановимо перемикач у положення «Новий робочий лист». Після завершення параметрів натисніть кнопку (ОК). Кінцевий екран налаштувань має виглядати так:

Результати дисперсійного аналізу будуть представлені на новому аркуші та складаються з двох таблиць. У першій таблиці для кожного рядка та кожного стовпця вихідної таблиці наведено числові параметри: кількість чисел, сума, середня та дисперсія.

	A	B	C	D	E	F	G
1	Двухфакторный дисперсионный анализ без повторений						
2							
3	ИТОГИ	Счет	Сумма	Среднее	Дисперсия		
4	Строка 1	6	68	11,33333	2,26666667		
5	Строка 2	6	79	13,16667	8,56666667		
6	Строка 3	6	88	14,66667	7,46666667		
7							
8	Столбец 1	3	35	11,66667	4,33333333		
9	Столбец 2	3	36	12	1		
10	Столбец 3	3	43	14,33333	4,33333333		
11	Столбец 4	3	42	14	7		
12	Столбец 5	3	49	16,33333	6,33333333		
13	Столбец 6	3	30	10	1		
14							
15							
16	Дисперсионный анализ						
17	Источник вариации	SS	df	MS	F	P-Значение	F критическое
18	Строки	33,44444	2	16,72222	11,4885496	0,002564116	4,102821015
19	Столбцы	76,94444	5	15,38889	10,5725191	0,000965646	3,32583453
20	Погрешность	14,55556	10	1,455556			
21							
22	Итого	124,9444	17				

Проаналізуємо отримані результати. Відповідно до другої таблиці маємо за строчками $F = 11,489$ і $F_{кр.} = 4,103$, тобто $F > F_{кр.}$. Відповідно, за стовпчиками, $F = 10,573$ і $F_{кр.} = 3,326$, тобто $F > F_{кр.}$. Отже вплив чинників на результативну ознаку вважатиметься достовірним, що й підтверджується величинами значимості ($p = 0,0026$ та $p = 0,00097$), що менше $0,05$.

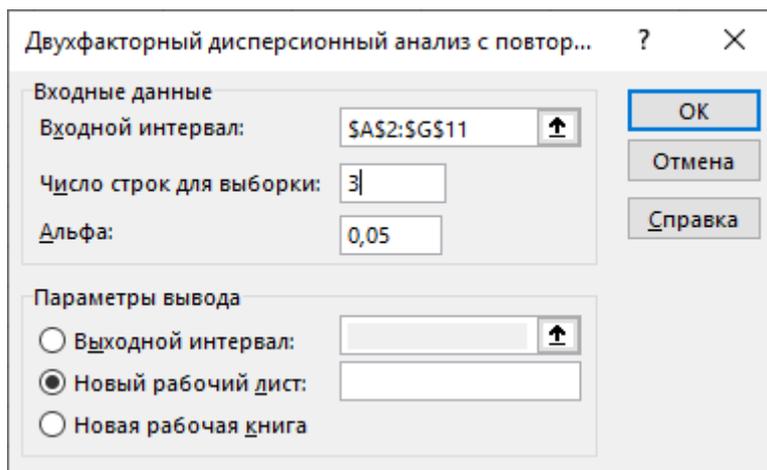
Таким чином, альтернативна гіпотеза про достовірність впливу факторів форми економічного укладу європейських країн XIV століття та географічного розміщення цих країн на смертність їхніх жителів під час пандемії може бути прийнята з ймовірністю 0.95 . При цьому сила впливу цих факторів оцінюється так:

$$K_A = \frac{110,389}{124,944} * 100\% = 88,4\%.$$

Змінімо умову попередньої задачі. Знову необхідно методом дисперсійного аналізу оцінити вплив форми економічного укладу (1 – аграрний, 2 – змішаний, 3 – промисловий) європейських країн XIV століття та географічного розміщення цих країн на смертність їхніх жителів під час пандемії. Відповідно до таблиці фактор А має 3 рівні (1–3), а фактор В – 6 рівнів (країни 1–6). Але цього разу в першій групі буде по 3 спостереження для кожного рівня (говорячи мовою історії – існує 3 джерела різної інформації щодо особливостей впливу фактору форми економічного укладу європейських країн XIV століття на рівень смертності їхніх жителів під час пандемії). Тому будемо використовувати двофакторний дисперсійний аналіз з повтореннями. Вихідні дані представлені у таблиці.

	A	B	C	D	E	F	G
1	смертність (%)						
2	фактори	країна 1	країна 2	країна 3	країна 4	країна 5	країна 6
3	1	10	11	12	11	14	10
4		7	8	10	9	11	11
5		13	14	15	14	18	13
6	2	11	12	15	16	16	9
7		10	12	13	17	15	7
8		5	6	9	9	9	5
9	3	14	13	16	15	19	11
10		6	5	8	7	9	6
11		10	9	12	12	16	8
12							

Виконуємо команду *Аналіз даних* із меню *Дані*. Відповідно до умов завдання вибираємо у діалоговому вікні метод «*Двофакторний дисперсійний аналіз з повтореннями*» і натискаємо кнопку (ОК). У вікні «*Двофакторний дисперсійний аналіз без повторень*» встановимо для вхідних даних такі параметри: вхідний інтервал дорівнює $A\$2:\$G\$11$, число строчок для вибірки 3, вхідний діапазон не містить мітки, альфа (рівень значимості) дорівнює 0,05. Для параметрів виводу встановимо перемикач у положення «*Новий робочий лист*». Після завершення параметрів натисніть кнопку (ОК). Кінцевий екран налаштувань має виглядати так:



Результати дисперсійного аналізу будуть представлені на новому аркуші та складаються з п'яти таблиць.

	A	B	C	D	E	F	G	H
1	Двухфакторный дисперсионный анализ с повторениями							
2								
3	ИТОГИ	країна 1	країна 2	країна 3	країна 4	країна 5	країна 6	Итого
4	1							
5	Счет	3	3	3	3	3	3	18
6	Сумма	30	33	37	34	43	34	211
7	Среднее	10	11	12,33333	11,33333	14,33333333	11,33333333	11,72222
8	Дисперсия	9	9	6,333333	6,333333	12,33333333	2,33333333	7,271242
9								
10	2							
11	Счет	3	3	3	3	3	3	18
12	Сумма	26	30	37	42	40	21	196
13	Среднее	8,666667	10	12,33333	14	13,33333333	7	10,88889
14	Дисперсия	10,33333	12	9,333333	19	14,33333333	4	14,9281
15								
16	3							
17	Счет	3	3	3	3	3	3	18
18	Сумма	30	27	36	34	44	25	196
19	Среднее	10	9	12	11,33333	14,66666667	8,33333333	10,88889
20	Дисперсия	16	16	16	16,33333	26,33333333	6,33333333	16,10458
21								
22	Итого							
23	Счет	9	9	9	9	9	9	
24	Сумма	86	90	110	110	127	80	
25	Среднее	9,555556	10	12,22222	12,22222	14,11111111	8,88888889	
26	Дисперсия	9,277778	10	7,944444	12,19444	13,61111111	6,86111111	

У перших чотирьох таблицях для кожного рядка та кожного стовця вихідної таблиці та для підсумкових їх значень наведено числові параметри: кількість чисел, сума, середнє та дисперсія. В

останній таблиці наведені результати двофакторного дисперсійного аналізу з повтореннями.

	A	B	C	D	E	F	G
28							
29	Дисперсионный анализ						
30	<i>Источник вариации</i>	<i>SS</i>	<i>df</i>	<i>MS</i>	<i>F</i>	<i>P-Значение</i>	<i>F критическое</i>
31	Выборка	8,333333	2	4,166667	0,35489	0,703676611	3,259446306
32	Столбцы	180,3889	5	36,07778	3,072871	0,02068723	2,477168673
33	Взаимодействие	48,11111	10	4,811111	0,409779	0,932934727	2,10605391
34	Внутри	422,6667	36	11,74074			
35							
36	Итого	659,5	53				
37							

Проаналізуємо отримані результати. Відповідно до останньої таблиці маємо за строчками $F = 0,35489$ і $F_{кр.} = 3,259$, тобто $F < F_{кр.}$. Відповідно, за стовпчиками, $F = 3,073$ і $F_{кр.} = 2,477$, тобто $F > F_{кр.}$. За показниками взаємодії факторів маємо $F = 0,41$ і $F_{кр.} = 2,106$, тобто $F < F_{кр.}$. Отже вплив чинника (географічного розміщення країн) на результативну ознаку (смертність їхніх жителів під час пандемії) вважатиметься достовірним, що й підтверджується величиною значимості ($p = 0,0207$), що менше $0,05$. При цьому сила впливу цього фактору є порівняно слабкою і оцінюється так:

$$K_A = \frac{180,39}{659,5} * 100\% = 27,35\%.$$

Щодо впливу чинника «форма економічного укладу (1 – аграрний, 2 – змішаний, 3 – промисловий) європейських країн XIV століття» на результативну ознаку (смертність їхніх жителів під час пандемії) та впливу на результативну ознаку взаємодії обох чинників приймаємо нульову гіпотезу щодо відсутності впливу названих факторів на результативну ознаку, так як в обох випадках $F < F_{кр.}$.

Таким чином, альтернативна гіпотеза про достовірність впливу факторів форми економічного укладу європейських країн XIV століття та географічного розміщення цих країн на смертність їхніх жителів під час пандемії за результатами розгляду трьох джерел даних може бути прийнята з ймовірністю $0,95$ лише у випадку впливу географічного розміщення країн на смертність їхніх жителів під час пандемії.

ЗАКЛЮЧЕННЯ

Таким чином, можемо констатувати, що застосування математико-статистичних методів є важливим етапом модернізації історичної науки. Вони не лише поглиблюють розуміння минулого, а й сприяють створенню інноваційних підходів до вивчення складних історичних явищ.

По-перше, математико-статистичні методи розширюють аналітичний інструментарій історика, дозволяють аналізувати великі обсяги історичних даних, встановлювати закономірності, визначати тренди й моделювати історичні процеси. Вони доповнюють традиційні якісні підходи, створюючи можливість для більш об'єктивних і детальних висновків.

По-друге, математико-статистичні методи уможливають використання кількісних даних для підтвердження гіпотез. Історичні джерела часто містять числову інформацію, яка може бути інтерпретована з використанням статистичних методів. Це сприяє підтвердженню або спростуванню історичних гіпотез та надає дослідженням більшої аргументованості.

По-третє, математико-статистичні методи дають можливість історикам застосовувати міждисциплінарний підхід у свої дослідженнях. Застосування математичних та статистичних методів сприяє міждисциплінарній взаємодії історії з іншими галузями знань, зокрема з економікою, демографією, соціологією та інформатикою. Такий підхід дозволяє глибше зрозуміти комплексні історичні явища.

По-четверте, математико-статистичні методи уможливають якісну візуалізацію даних, адже завдяки графікам, діаграмам, картам і мережам дослідники можуть краще унаочнювати

історичні процеси. Це полегшує інтерпретацію даних і сприяє комунікації результатів досліджень широкій аудиторії.

Втім, при застосуванні таких методів можуть бути потенційні ризики і обмеження. Це пов'язано з тим, що хоча математико-статистичні методи й відкривають нові горизонти для історичних досліджень, важливо уникати їхнього механістичного застосування. Дані мають бути ретельно відібрані, а методи – адаптовані до специфіки історичних явищ. Неправильна інтерпретація або спрощення можуть призвести до викривлення історичних реалій.

Крім того, застосування математико-статистичних методів має виражені етичні аспекти. Аналіз історичних даних, особливо пов'язаних із демографічними, соціальними чи конфліктними питаннями, вимагає відповідального підходу. Дослідники повинні враховувати культурний і соціальний контекст, аби уникати некоректних або упереджених висновків.

Але безперечним є факт, що застосування математико-статистичних методів в гуманітарних дослідженнях має перспективи розвитку. Інтеграція сучасних технологій, таких як машинне навчання, аналіз «великих даних» (Big Data) і геоінформаційні системи (ГІС), значно розширює можливості математико-статистичних методів в історичних і, загалом, гуманітарних, дослідженнях. Це створює нові можливості для вивчення історії, які раніше були недоступними. Математико-статистичні методи є потужним інструментом у сучасних історичних дослідженнях, що дозволяє по-новому осмислювати минуле, здійснювати аналіз великих масивів даних та робити обґрунтовані висновки. У процесі створення цього посібника було продемонстровано, як кількісні методи, зокрема статистичні аналізи, математичне моделювання, візуалізація даних і просторовий аналіз, можуть інтегруватися в дослідницьку практику істориків.

Особливу увагу приділено важливості коректного використання математичних і статистичних методів, що включає вибір відповідних інструментів, інтерпретацію отриманих

результатів у контексті історичних реалій та уникнення механістичного підходу.

Посібник може стати корисним для студентів, аспірантів, викладачів і дослідників, які прагнуть розширити свої методологічні горизонти та вийти за межі традиційного наративного підходу.

Закликаємо читачів активно використовувати описані методи у своїх дослідженнях, адаптуючи їх до специфіки власних тем і джерел. Інтеграція математико-статистичних підходів з іншими методами історичної науки сприятиме побудові більш комплексного та глибокого розуміння минулого, а також розвитку міждисциплінарної співпраці між гуманітарними й природничими науками.

Ми сподіваємося, що цей посібник стане для Вас не лише джерелом нових знань, але й стимулом для подальших відкриттів у сфері історичної науки.

ЛІТЕРАТУРА

1. A Million Random Digits with 100,000 Normal Deviates. Originally published: Glencoe, Ill.: Free Press, 1955.
2. Akbash, K.S. & Pasichnyk, N.O. & Rizhniak, R.Ya. (2021). Analysis of key factors of influence on scientometric indicators of higher educational institutions of Ukraine. International Journal of Educational Development, Elsevier, vol. 81.
3. Atteslander P. Methoden der empirischen Sozialforschung. 10. Auflage. Berlin, 2003. S. 215–249.
4. FrühW. Inhaltsanalyse: Theorie und Praxis. 6. Auflage. Berlin, 2003. S. 27.
5. Gender-sensitive Education Statistics and Indicators a practical Guide. United Nations Educational, Scientific and Cultural Organization. 1997. 37 p.
6. Global Gender Gap Report 2006. URL: <https://www.weforum.org/reports/global-gender-gap-report-2006>
7. Human Development Report 1990. New York, Oxford: Oxford University Press, 1990. URL: http://hdr.undp.org/sites/default/files/reports/219/hdr_1990_en_complete_nostats.pdf
8. Human Development Report 2010. URL: http://hdr.undp.org/sites/default/files/reports/270/hdr_2010_en_complete_reprint.pdf
9. Human Development Report. 1995. New York. URL: http://hdr.undp.org/sites/default/files/reports/256/hdr_1995_en_complete_nostats.pdf

10. Natalya Pasichnyk, Renat Rizhniak, Hanna Deforzsh. Biographical Materials of Mathematicians and Natural Scientists in “The Bulletin of Experimental Physics and Elementary Mathematics” (1886–1917): meaningful and content analysis. *History of science and technology*, 2022, vol. 12, issue 2. 279–301. <https://doi.org/10.32703/2415-7422-2022-12-2-279-301>
11. Pasichnyk, N., Rizhniak, R., Deforzsh, H. Congresses of natural scientists and mathematicians in the “Bulletin of experimental physics and elementary mathematics” (1886–1917): Analysis of publications. *History of Science and Technology*, 2023. 13 (2), 280–310. <https://doi.org/10.32703/2415-7422-2023-13-2-280-310>
12. Patrimonium. Студії з ранньомодерної історії Центрально-Східної Європи. Т. I: Ранньомодерна людина: Простір – влада – право XVI–XVIII століть / за ред. В. Михайловського і Я. Столицького. Київ – Краків, Laurus – Historia Iagellonica, 2015. 408 с.
13. Technical notes. Calculating the human development indices – graphical presentation. Human Development Indices and Indicators. Statistical Update. 2018. URL: http://www.hdr.undp.org/sites/default/files/hdr2018_technical_notes.pdf
14. The Global Gender Gap Report 2016
15. Women and men 17 in Kirovohrad region. Statistical book. // Кропивницький. 2016. 104 р.
16. Акбаш К.С., Пасічник Н.О., Ріжняк Р.Я. Відображення особливостей освітнього, наукового й соціального середовища в текстах місій закладів вищої освіти України. Аналітичний звіт / Центральноукраїнський державний педагогічний університет імені Володимира Винниченка, Міждисциплінарний науковий центр прикладних досліджень. Кропивницький, 2021. 124 с. URL: <http://surl.li/rzgdhe>
17. Акбаш К.С., Пасічник Н.О., Ріжняк Р.Я. Прогностичний аналіз публікаційної активності науковців університетів України у контексті їх входження до рейтингу QS World University

- Rankings. *Статистика України*. 2020. 4. 61–71.
[http://doi.10.31767/su.4\(91\)2020.04.07](http://doi.10.31767/su.4(91)2020.04.07)
18. Акбаш К.С., Пасічник Н.О., Ріжняк Р.Я. Публікаційна активність учених педагогічних університетів України (2010–2020 рр.) через призму баз даних Scopus та Web of Science. *Наука та наукознавство*. 2021. 1 (111). 63–80.
<https://doi.org/10.15407/sofs2021.01.063>
19. Бевз Т.А. Феномен «революція» у дискурсах мислителів, політиків, науковців. Монографія. Київ, 2012. 176 с.
20. Гербут Н.А. Теоретико-методологічні засади гендерних досліджень. *Наукові записки Інституту політичних і етнонаціональних досліджень ім. І.Ф. Кураса*. 2012. Вип. 1. С. 344–361.
21. Еволюція міграційних процесів крізь призму теорії міжнародних відносин. *Науковий вісник Ужгородського національного університету*. 24 (1). 2019. 10–13/
22. Історія міжнародних економічних відносин: Курс лекцій / І.М. Грабинський, І.В. Михайляк, Л.А. Українець та ін. За ред. проф. І.М.Грабинського / ЛНУ ім. Івана Франка. Львів, НВФ «Українські технології», 2013. 340 с.
23. Коваліско Н.В. Основи соціальної стратифікації: Навчальний посібник. Львів, «Магнолія 2006», 2009. 328с.
24. Куліков В.О. Нові інформаційні технології в історичних дослідженнях. *Вісник Харківського національного університету ім. В.Н. Каразіна. Серія Історія*. 2007. № 762. С. 217–228.
25. Міжнародний банк даних <https://data.worldbank.org/indicator>
26. Пасічник Н.О., Ріжняк Р.Я. Е.К. Шпачинський – засновник науково-популярної журналістики кінця ХІХ ст. на теренах України. *Історія науки і біографістика*. 2024. 1. 91–123.
<https://doi.org/10.31073/istnauka202401-07>
27. Підгорний А.З., Самоєнкова О.В., Ольвінська Ю.О., Вітковська К.В. Соціально-демографічна статистика: Підручник. Одеса, ФОП Гуляєва В.М., 2016. 424 с.

28. Ріжняк Р.Я. Аналіз забезпечення комп'ютерною технікою технічних університетів України протягом 1991–2011 років. *Інформаційні технології і засоби навчання*. 2017. 60 (4). 46–57. <https://doi.org/10.33407/itlt.v60i4.1748>
29. Ріжняк Р.Я. Контент-аналіз нумізматичних досліджень українських історичних періодичних видань (1996–2016 роки). *Наукові праці історичного факультету Запорізького національного університету*. 2016. 45 (2). 103–107.
30. Орлик В. М., Орлик С. В.. Теоретико-методологічні та джерелознавчі проблеми економічної історії України. *The Universe of History and Archeology*. Дніпро, 2019. Т. 2(27). Вип. 2. С. 5-24. DOI: 10.15421/26190201.
31. Чума, холера, «іспанка»: як великі пандемії змінювали світ. URL: <https://www.bbc.com/ukrainian/features-52291698>
32. Ямелинець Т.С. Застосування географічних інформаційних систем у ґрунтознавстві: Навчальний посібник. Львів, Видавничий центр ЛНУ імені Івана Франка, 2008. 196 с.

МАТЕМАТИКО-СТАТИСТИЧНІ МЕТОДИ В ІСТОРИЧНИХ ДОСЛІДЖЕННЯХ

Навчальний посібник
для здобувачів вищої освіти освітньо-наукового рівня
галузі знань 03 Гуманітарні науки

Пасічник Наталя Олексіївна

професор кафедри математики та цифрових технологій
Центральноукраїнського державного університету
імені Володимира Винниченка,
доктор історичних наук, професор

Ріжняк Ренат Ярославович

професор кафедри математики та цифрових технологій
Центральноукраїнського державного університету
імені Володимира Винниченка,
доктор історичних наук, професор