

**НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ ЦИВІЛЬНОГО ЗАХИСТУ
УКРАЇНИ**

**Факультет пожежної безпеки
Кафедра пожежної профілактики в населених пунктах**

**МЕТОДИЧНІ ВКАЗІВКИ
ДО ПРАКТИЧНИХ ЗАНЯТЬ З НАВЧАЛЬНОЇ ДИСЦИПЛІНИ
«МЕТОДИ ОБРОБКИ КІЛЬКІСНИХ ПОКАЗНИКІВ ПОЖЕЖНОЇ
БЕЗПЕКИ»**

**для здобувачів третього (освітньо-наукового) рівня вищої освіти
галузь знань 26 «Цивільна безпека», спеціальність 261 «Пожежна
безпека»**

Харків 2023

Рекомендовано до друку
кафедрою пожежної
профілактики в населених
пунктах НУЦЗ України
(протокол № 7 від 27.02.2023 р.)

Укладачі: Н.В. Рашкевич, Ю.А. Отрош

Рецензенти: доктор технічних наук, професор **М. Г. Сур'янінов**,
завідувач кафедри будівельної механіки Одеської державної
академії будівництва та архітектури;
кандидат технічних наук, доцент **О. Б. Васильєв**,
начальник Дарницького районного управління Головного
управління Державної служби України з надзвичайних
ситуацій у м. Києві

Методичні вказівки до практичних занять з навчальної дисципліни
«Методи обробки кількісних показників пожежної безпеки» для здобувачів
третього (освітньо-наукового) рівня вищої освіти галузь знань 26
«Цивільна безпека», спеціальність 261 «Пожежна безпека» / Укладачі:
Н.В. Рашкевич, Ю.А. Отрош. Х.: НУЦЗУ, 2023. 89 с.

ЗМІСТ

Вступ	4
Практичне заняття № 1 Експеримент як метод дослідження об'єктів	6
Практичне заняття № 2 Похибки вимірювань	11
Практичне заняття № 3 Класифікація даних та способи їх отримання	21
Практичне заняття № 4 Звітність та публікація результатів експерименту	26
Практичне заняття № 5 Представлення результатів обробки даних	31
Практичне заняття № 6 Представлення результатів обробки даних	38
Практичне заняття № 7 Однофакторний дисперсійний аналіз	43
Практичне заняття № 8 Двофакторний дисперсійний аналіз	53
Практичне заняття № 9 Побудова лінійної моделі. Загальні відомості	58
Практичне заняття № 10 Побудова лінійної моделі. Загальні відомості	65
Практичне заняття № 11 Побудова нелінійної моделі. Загальні відомості	76
Практичне заняття № 12 Модульна контрольна робота	87

ВСТУП

Мета навчальної дисципліни є формування у здобувачів вищої освіти системи базових знань і навичок для організації, проведення експериментальних досліджень та обробки їх результатів при дослідженні кількісних показників пожежної небезпеки, а також систематизація, розширення та закріплення професійних знань і формування навичок самостійного ведення наукової роботи, дослідження та експериментування.

Завдання навчальної дисципліни є ознайомлення з експериментальними методами проведення наукових досліджень та обробки їх результатів, розвиток практичних навичок з планування, організації та проведення експериментальних наукових досліджень, освоєння різних методів аналізу та обробки результатів експериментальних наукових досліджень, ознайомлення з методами побудови однофакторних, багатфакторних моделей та перевірки їх на адекватність, з методами пошуку та дослідження зв'язків між експериментальними даними, а також отримання практичних навичок з використання комп'ютерної техніки для обробки результатів експериментальних наукових досліджень.

У результаті вивчення навчальної дисципліни здобувач вищої освіти повинен отримати:

знання:

- методи планування та проведення експериментальних наукових досліджень для визначення кількісних показників пожежної небезпеки;
- методи обробки результатів експериментальних наукових досліджень;
- методи побудови однофакторних, багатфакторних моделей та перевірки їх на адекватність.

уміння/навички:

- застосовувати отримані навички для самостійного планування, підготовки і проведення експериментальних наукових досліджень;
- самостійно проводити обробку експериментальних даних при визначенні кількісних показників пожежної небезпеки;
- самостійно будувати однофакторні, багатфакторні моделі та перевіряти їх на адекватність за критеріями Ст'юдента, Фішера, Пірсона;
- використовувати комп'ютерну техніку для обробки результатів експериментальних наукових досліджень;
- кількісних показників пожежної небезпеки об'єктів різного призначення;

– розробки і обґрунтування заходів з посилення пожежної безпеки об'єктів об'єктах різного призначення.

комунікації:

– донесення до фахівців і нефахівців інформації, ідей, проблем, рішень, власного досвіду та аргументації;

– вільне спілкування з питань, що стосуються сфери наукових та експертних знань, з колегами, широкою науковою спільнотою, суспільством у цілому.

відповідальність та автономію:

– демонстрація значної авторитетності, інноваційність, високий ступінь самостійності, академічна та професійна доброчесність, послідовна відданість розвитку нових ідей або процесів у передових контекстах професійної та наукової діяльності;

– відповідальність за внесок до професійних знань і практики;

– здатність продовжувати навчання з високим ступенем автономії;

– здатність до безперервного саморозвитку та самовдосконалення.

ПРАКТИЧНЕ ЗАНЯТТЯ № 1

ЕКСПЕРИМЕНТ ЯК МЕТОД ДОСЛІДЖЕННЯ

Література:

1. Карагодова О.О., Рожок В.Д. Дослідження операцій: навч. посіб. К.: Центр учбової літератури, 2007. 256 с.
2. Бахрушин В.Є. Методи аналізу даних: навчальний посібник для студентів. Запоріжжя: КПУ, 2011. 268 с
3. Перегуда О.В., Капустян О.А., Курилко О.Б. Статистична обробка даних: навч. посіб. Електронне видання, 2022. 103 с.
4. Рогальський Ф. Б., Курилович Я. Є., Цокурєнко А. А. Математичні методи аналізу економічних систем. Книга 1. Київ: Наукова думка, 2001, 435 с.
5. Рогальський Ф. Б., Цокурєнка А. А. Математичні методи аналізу економічних систем. Книга 2. Київ: Наукова думка, 2001, 423 с.
6. Крайчук О.В., Московська Г.К., Соколенко О.І. Теорія ймовірностей і математична статистика. Рівне, 2004.

Мета:

1. Навчальна – вивчити основи організації та проведення експериментальних досліджень.
2. Виховна – виховувати професійні риси, почуття відповідальності; виховувати зацікавленість дисципліною, прагнення отримати нові знання самостійно.
3. Розвиваюча – придбання нових знань щодо основ експериментальних досліджень з метою вирішення питань у сфері пожежної безпеки.

Питання для обговорення:

1. Експеримент як метод дослідження.
2. Розробка плану - програми й методики проведення експерименту.

Теоретична частина:

Експеримент – апробація знання досліджуваних явищ в контрольованих або штучно створених умовах. Сам термін «експеримент» (від латинського – спроба, дослід) означає науково поставлений дослід, спостереження досліджуваного явища у певних умовах, що дозволяють багаторазово відтворювати його при повторенні цих умов.

Експеримент – важливий елемент наукової практики, вважається основою теоретичного знання, критерієм його дійсності.

Експеримент – це система операцій, впливу або спостережень, спрямованих на одержання інформації про об'єкт при дослідницьких випробуваннях, які можуть проводитись в природних і штучних умовах при зміні характеру проходження процесу.

Експеримент проводять на заключному етапі дослідження і він є критерієм істини теорії і гіпотез. Експеримент також у багатьох випадках є джерелом нових теоретичних даних, які розвиваються на базі результатів проведеного досліду або законів, що виходять з нього. Основною метою експерименту є перевірка теоретичних положень (підтвердження робочої гіпотези), а також більш широкого і глибокого вивчення теми наукового дослідження.

Експеримент – це спосіб дослідження явищ, процесів шляхом організації спеціальних дослідів, які забезпечують вивчення впливу окремих чинників за умови постійності інших умов або моделювання явищ і процесів на практиці. Експеримент має бути проведений за можливістю в короткі терміни з мінімальними затратами і з високою якістю отриманих результатів.

В методологічному відношенні експеримент передбачає перехід дослідника від пасивного до активного способу діяльності. Експеримент проводять у таких випадках:

- у разі необхідності відшукати в об'єкта раніше невідомі властивості;
- у разі перевірки правильності теоретичних викладок;
- у разі демонстрації явища.

Експеримент неможливий без теоретичних положень, які він або підтверджує, або спростовує, тому є одним із найважливіших шляхів розвитку сучасної науки.

Специфіка експериментального дослідження в науково-дослідній практиці має такі етапи:

- 1) констатуючий;
- 2) формуючий;
- 3) корегуючий;
- 4) контрольний.

Експеримент як науковий метод необхідний для спеціального вивчення окремих питань з участю тих, хто підлягає експерименту. Його реалізація передбачає створення спеціальних умов та груп за певними параметрами, яких потребує гіпотеза. Завдяки експерименту апробують навчальні програми з різних мистецьких дисциплін, з'ясовують їх ефективність.

Розрізняють:

- 1) експерименти, що перевіряють вірність гіпотези чи теорії емпірично;
- 2) експерименти пошукового плану, у ході яких відбувається відбір необхідної емпіричної інформації для уточнення запропонованого.

2. Розробка плану - програми й методики проведення експерименту.

План-програма включає найменування теми дослідження, робочу гіпотезу, методику експерименту, перелік необхідних матеріалів, приладів, установок, список виконавців експерименту, календарний план робіт і кошторис на виконання експерименту. У ряді випадків включають роботи з конструювання й виготовлення приладів, апаратів, пристосувань.

Один з найбільш важливих етапів складання плану-програми – **визначення мети й завдань експерименту**. Чітко обґрунтовані завдання - це вагомий внесок у їхнє рішення. Кількість завдань повинне бути невеликим. Для конкретного експерименту оптимальною кількістю є 3-4 завдання. У великому, комплексному експерименті їх може бути 8-10.

Необхідно правильно **вибрати фактори, що** варіюють, тобто встановити основні й другорядні характеристики, що впливають на досліджуваний процес. Спочатку аналізують розрахункові (теоретичні) схеми процесу. На основі цього класифікують всі фактори й становлять із них убутний по важливості для даного експерименту ряд. Правильний вибір основних і другорядних факторів відіграє важливу роль в ефективності експерименту, оскільки експеримент і зводиться до знаходження залежностей між цими факторами. У тих випадках, коли важко відразу виявити роль основних і другорядних факторів, виконують невеликий за обсягом пошуковий експеримент.

Обґрунтування засобів вимірів – це вибір необхідних для спостережень і вимірів приладів, устаткування, машин, апаратів й ін. Засоби виміру можуть бути обрані стандартні або у випадку відсутності таких - виготовлені самостійно.

Дуже відповідальною частиною є встановлення точності вимірів і погрішностей. Методи вимірів повинні базуватися на законах спеціальної науки – **метрології**.

У **методиці** докладно проектують процес проведення експерименту. На початку становлять послідовність (черговість) проведення операцій вимірів і спостережень. Потім ретельно описують кожну операцію окремо з урахуванням обраних засобів для проведення експерименту. Особливу увагу приділяють методам контролю якості операцій, що забезпечують при мінімальному (раніше встановленому) кількості вимірів високу надійність і задану точність.

На обсяг і трудомісткість істотно впливає вид експерименту. Полеві експерименти, як правило, мають більшу трудомісткість.

Проведення експерименту є найважливішим і найбільш трудомістким етапом. Експериментальні дослідження необхідно проводити відповідно до затвердженої плану-програми й особливо методикою експерименту. Приступаючи до експерименту, остаточно уточнюють методику його проведення, послідовність випробувань.

При складному експерименті часто виникають випадки, коли очікуваний результат одержують пізніше, ніж передбачається планом. Тому науковець повинен виявити терпіння, витримку, наполегливість і довести експеримент до одержання результатів.

Особливе значення має сумлінність при проведенні експериментальних робіт; неприпустима недбалість, що приводить до більших перекручувань, помилок. Порушення цих вимог - до повторних експериментів, що продовжує дослідження.

Обов'язковою вимогою проведення експерименту є **ведення журналу**. Форма журналу може бути довільною, але повинна щонайкраще відповідати досліджуваному процесу з максимальною фіксацією всіх факторів. У журналі відзначають тему НДР і тему експерименту, прізвище виконавця, час і місце проведення експерименту, характеристику навколишнього середовища, дані про об'єкт експерименту й засобах виміру, результати спостережень, а також інші дані для оцінки одержуваних результатів.

Журнал потрібно заповнювати акуратно, без будь-яких виправлень. При одержанні в одному статистичному ряді результатів, що різко відрізняються від сусідніх вимірів, виконавець повинен записати всі дані без перекручувань і вказати обставини, що супроводжують зазначеному виміру. Це дозволить установити причини перекручувань і кваліфікувати виміру як відповідному реальному ходу процесу або як грубий промах.

Одночасно з вимірами виконавець повинен проводити попередню обробку результатів й їхній аналіз. Тут особливо повинні проявлятися його творчі здатності. Такий аналіз дозволяє контролювати досліджуваний процес, коректувати експеримент, поліпшувати методику й підвищувати ефективність експерименту.

Важливі при цьому консультації з колегами по роботі й особливо з науковим керівником. У процесі експерименту необхідно дотримуватись вимоги інструкцій з промсанітарії, техніці безпеки, пожежній профілактиці.

Спочатку результати вимірів зводять у таблиці по характеристиках, що варіюють, для різних досліджуваних питань. Дуже ретельно уточнюють сумнівні цифри. Установлюють точність обробки досвідчених даних.

Важливим розділом методики є **вибір методів обробки й аналізу експериментальних даних**. Обробка даних зводиться до систематизації всіх цифр, класифікації, аналізу. Результати експериментів повинні бути зведені в зручні по формі записи - таблиці, графіки, формули, номограми, що дозволяють швидко й доброякісно зіставляти отримані результати.

Особлива увага в методиці повинна бути приділена **математичним методам обробки й аналізу даних** - установленню емпіричних залежностей, апроксимації зв'язків між характеристиками, що варіюють, установленню критеріїв і довірчих інтервалів й ін.

Теми доповідей для заслуховування:

1. Сутність, мета, функції наукового експерименту.
2. Класифікація експериментів.
3. Методологія експериментальних досліджень.
4. Загальні вимоги до проведення експерименту.
5. Типові помилки в проведенні експерименту.
6. Робоче місце експериментатора та організація експерименту.

Питання для контролю знань:

1. Надайте визначення: «експеримент», «дослід», «спостереження», «методика експерименту».
2. Мета експерименту.
3. Ознаки класифікації експерименту.
4. Випадки проведення експерименту.
5. Етапи підготовки до проведення експериментального дослідження.
6. Обґрунтування засобів вимірів.
7. Умови розробки методики експерименту.
8. У чому полягає обробка і аналіз експериментальних даних.

Завдання для самостійної підготовки:

I. Вивчити матеріал лекції.

II. Підготувати доповідь за темою:

1. Класифікація видів та методів вимірювання.
2. Метрологічні характеристики засобів вимірювання.
3. Основи теорії похибок вимірювань.
4. Міжнародна система одиниць фізичних величин.
5. Класифікація та характеристики похибок.
6. Розрахунок похибок при прямих вимірах.
7. Розрахунок похибок при непрямих вимірах.
8. Систематичні похибки та їх корекція.

III. Пройти тестування: тест 1.2. за посиланням
<https://forms.gle/3bR8hwMhTy8NGAuh6>

ПРАКТИЧНЕ ЗАНЯТТЯ № 2 ПОХИБКИ ВИМІРЮВАНЬ

Література:

1. Карагодова О.О., Рожок В.Д. Дослідження операцій: навч. посіб. К.: Центр учбової літератури, 2007. 256 с.
2. Бахрушин В.Є. Методи аналізу даних: навчальний посібник для студентів. Запоріжжя: КПУ, 2011. 268 с
3. Перегуда О.В., Капустян О.А., Курилко О.Б. Статистична обробка даних: навч. посіб. Електронне видання, 2022. 103 с.
4. Рогальський Ф. Б., Курилович Я. Є., Цокурєнко А. А. Математичні методи аналізу економічних систем. Книга 1. Київ: Наукова думка, 2001, 435 с.
5. Рогальський Ф. Б., Цокурєнка А. А. Математичні методи аналізу економічних систем. Книга 2. Київ: Наукова думка, 2001, 423 с.
6. Крайчук О.В., Московська Г.К., Соколенко О.І. Теорія ймовірностей і математична статистика. Рівне, 2004.

Мета:

1. Навчальна – вивчити основи експериментальних досліджень, теорію похибок.
2. Виховна – виховувати професійні риси, почуття відповідальності; виховувати зацікавленість дисципліною, прагнення отримати нові знання самостійно.
3. Розвиваюча – придбання нових знань з теорії похибок.

Питання для обговорення:

1. Фізичні величини та їх вимірювання.
2. Похибки вимірювання.
- 2.1. Класифікація похибок вимірювань.

Теоретична частина:

1. Фізичні величини та їх вимірювання.

Вимірювання – сукупність переважно експериментальних операцій, що виконуються за допомогою технічного засобу, що зберігає одиницю величини, що дозволяє зіставити вимірювану величину з її одиницею і отримати шукане значення величини. Це значення називають результатом вимірювання.

Для встановлення відмінності в кількісному значенні відображуваного об'єкта введено поняття фізичної величини.

Фізичною величиною (ФВ) називається одна з властивостей фізичного об'єкта (явища, процесу), загальне в якісному відношенні для багатьох фізичних об'єктів, але в кількісному відношенні індивідуальне для кожного об'єкта

Поняття фізичної величини (ФВ) – одне з найбільш загальних понять у фізиці та метрології. Під ФВ слід розуміти властивість, загальну в якісному відношенні для безлічі матеріальних об'єктів, але індивідуальну в кількісному відношенні для кожного з них. Так усі об'єкти мають масу і температуру, але для кожного конкретного об'єкта як маса, так і температура різні залежно від різних обставин.

Для встановлення різниці в кількісному відношенні властивостей кожного об'єкта вводиться поняття «розмір ФВ».

Між розмірами кожної ФВ існують відносини, які мають таку ж саму логічну структуру, що й між числовими формами (цілими, раціональними, дійсними числами, векторами). Тому численність числових форм з відносинами між ними може слугувати моделлю ФВ, тобто безлічі її розмірів і відносин між ними.

Можна виділити три групи ФВ, вимірювання яких проводиться за принципово різними правилами.

До першої групи належать величини, безліч розмірів яких визначаються за відношенням типу «твердий - м'який», «тепле - холодне», «кисле - солодке» і т. д. У математиці такі відносини отримали назву «відносини порядку та еквівалентності». Наявність подібних відносин встановлюється теоретично, виходячи із загально-фізичних міркувань, або експериментально, за допомогою засобів вимірювання й експериментатора. Так без особливих зусиль можна визначити, що мідь твердіша за гуму, але визначити відмінність міді за твердістю з іншими металами (свинцем, оловом) без засобів вимірювання неможливо, тому що їх твердість відрізняється незначно.

Друга група величин характеризується тим, що відношення порядку й еквівалентності стосується не тільки розмірів величин, а й відмінностей у парах цих величин (потенціал, енергія, температура та ін.). Так інтервал температур буде однаковим, якщо будуть однакові відстані між поділками на шкалі ртутного термометра. Мова йде не про температуру як ступінь нагрівання, а лише про рівність інтервалів температур.

До третьої групи величин, крім зазначених раніше визначень, відносяться характерні відносини, які називаються операціями, подібно до арифметичного додатка (множення на π) та вирахування. Результат відповідає сумі π розмірів певної вимірюваної величини. До таких величин відносяться: довжина, тиск, маса, термодинамічна температура і т. д. До величин третьої групи можна віднести і безліч інтервалів розмірів величин другої групи, тому що для них можливо встановити операцію, подібну до додатка. Отже, ці величини є найбільш зручними для використання. Тому їх і називають фізичними.

Між властивостями об'єкта існують взаємозв'язки. Модель об'єкта описується сукупністю рівнянь, які називаються рівняннями між величинами. У кожному розділі науки кількість рівнянь завжди менша, ніж кількість вхідних величин. Тому в окрему групу прийнято виділяти

величини, кількість яких дорівнює різниці між кількістю величин і кількістю незалежних рівнянь.

Ці величини і відповідні їм одиниці вимірювання називаються основними величинами і основними одиницями. Решта величин та одиниць, які однозначно визначаються через основні, називаються похідними.

ФВ поділяються на вимірювані й оцінювані.

Вимірювані ФВ можуть бути виражені кількісно у вигляді певного числа встановлених одиниць виміру. ФВ, для яких неможливо ввести одиницю виміру, можуть бути тільки оцінювальними (землетрус -7 балів, шторм - 10 балів, твердість за шкалою Мооса - 5 балів і т. д.).

За видами появи ФВ поділяються на 3 групи:

- речовинні (пасивні), тобто ті, що описують фізичні та фізико-хімічні властивості речовин, матеріалів і виробів з них. До цієї групи належить маса, щільність (питома вага), електричний опір, ємність, індуктивність та ін. Іноді наведені ФВ називають пасивними. Для їх вимірювань необхідно використовувати додаткові джерела енергії, за допомогою яких формується сигнал інформації, що вимірюється. При цьому пасивні ФВ перетворюються в активні, які й вимірюються;

- енергетичні (активні), тобто величини, які описують характеристики процесів перетворення, передачі і використання енергії. До них відносяться струм, напруга, потужність, енергія. Ці величини називають активними. Вони можуть бути перетворені на сигнали вимірювальної інформації без використання додаткових джерел енергії;

- які характеризують протікання процесів у часі. До цієї групи відносяться різного виду спектральні характеристики, кореляційні функції і т. ін.

За належністю до різних груп фізичних процесів ФВ поділяються на просторово-тимчасові, механічні, теплові і т. д.

За ступенем умовної незалежності від інших величин даної групи ФВ поділяються на основні (умовно незалежні), похідні (умовно залежні) та позасистемні.

За наявності розмірності ФВ поділяються на розмірні, тобто ті, які мають розмірність, і безрозмірні.

Розмір (фізичної) величини - кількісний вміст ФВ у даному об'єкті. Не слід використовувати термін «величина» як кількісну характеристику даної властивості, наприклад, у термінах «величина напруги», «величина маси» та ін. У таких випадках слід використовувати термін «розмір напруги», «розмір маси».

Основна (фізична) величина - ФВ, що входить до системи ФВ і прийнята за незалежну від інших величин цієї системи.

Похідна (фізична) величина - ФВ, що входить до системи величин та визначається через основні величини цієї системи.

Розмірність фізичних величин - вираз, що відображає її зв'язок з основними величинами системи величин.

Розмірність основної фізичної величини - умовний символ ФВ у даній системі величин.

Властивості – це певні характеристики, здатності, які притаманні тому чи іншому фізичному тілу.

Вимірювання – зводиться до порівняння певної властивості тіла з одиницею даної фізичної величини.

Фізична величина – властивість, спільна в якісному відношенні у багатьох матеріальних об'єктів та індивідуальна в кількісному відношенні у кожного з них.

Фізичний прилад – це пристрій чи засіб, який служить для вимірювання певних фізичних величин.

2. Похибки вимірювання.

Процедура вимірювання складається з таких основних етапів:

- прийняття моделі об'єкту вимірювання;
- вибір методу вимірювання;
- вибір засобів вимірювань;
- проведення розрахунків з метою обрання числового значення результату вимірювання.

Різного роду недоліки, властиві цим етапам, призводять до того, що результат вимірювання відрізняється від істинного значення вимірюваної величини. Величину, що характеризує відхилення результату вимірювання від істинного значення вимірюваної величини, називають похибкою вимірювання.

Похибки є властивістю будь-якого вимірювання, вони обумовлені пізнавальним характером процесу вимірювання відносно до наших знань. В науці слово “похибка” не має звичайного значення чогось невірною.

Похибки, власне, не слід відносити до помилок експериментатора, їх неможливо уникнути, намагаючись бути дуже уважним. Найкраще, на що можна розраховувати – це звести похибки до мінімуму і надійно розрахувати їх величини.

Причин, що призводять до появи похибок при вимірюваннях, надзвичайно багато. Вони зумовлені:

- недостатнім знанням властивостей досліджуваного об'єкта;
- недосконалістю методів та засобів вимірювань;
- зовнішніми впливами (тиску, вологості тощо);
- суб'єктивними властивостями оператора (експериментатора);
- округленням результатів вимірювання до парних чи непарних чисел тощо.

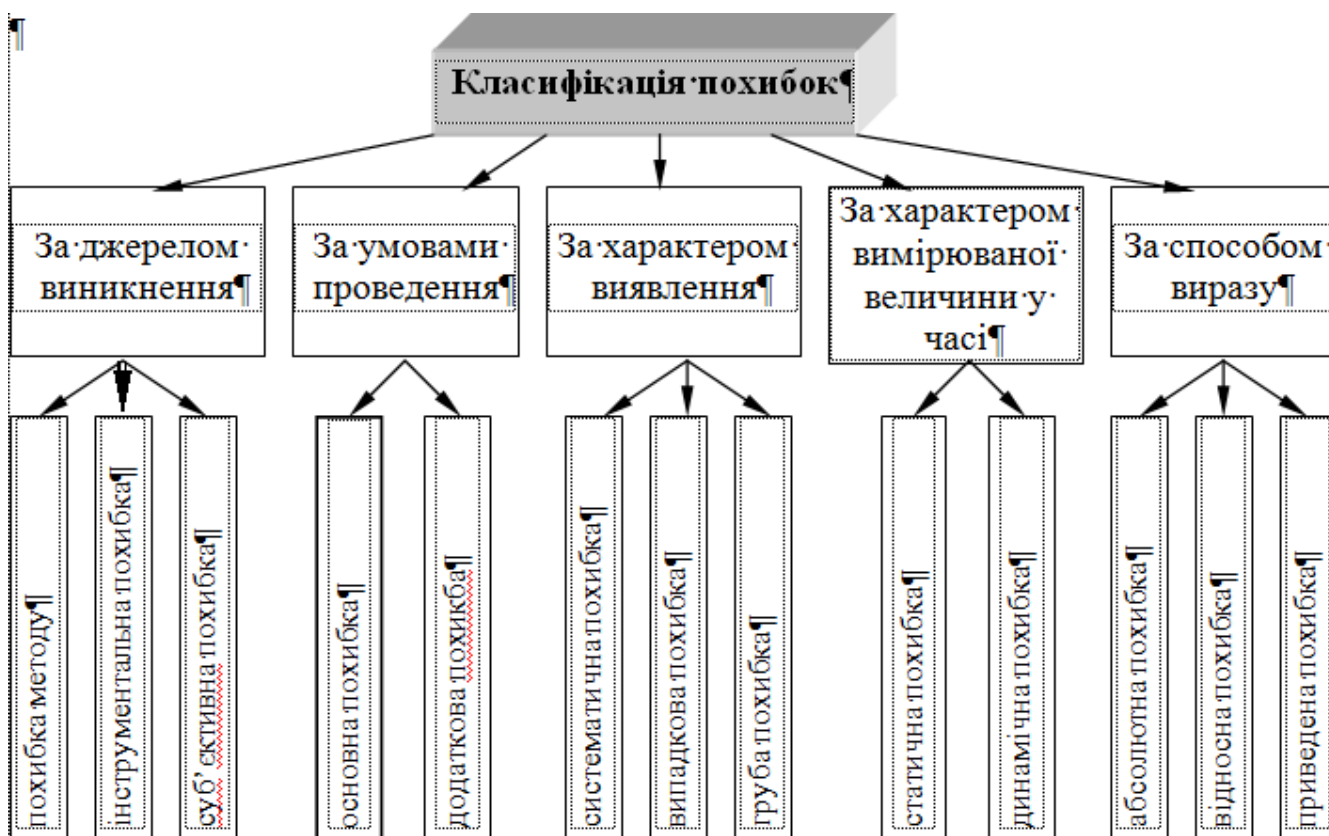
Звичайно, одним з основних завдань при проведенні будь-якого вимірювання є виявлення та усунення причин та завад, що призводять до появи похибок.

Можна зробити висновок, що жодне з вимірювань, як би ретельно воно не проводилося, не може обійтись без похибок. В зв'язку з тим, що істинне значення вимірюваної величини залишається невідомим, неможливо визначити й істинне значення похибки вимірювання, тобто на практиці можна знайти лише наближене значення похибки, виконати її оцінку.

Таким чином, до задачі вимірювання входить не тільки визначення значення ФВ, але також й оцінка похибки, що була допущена під час вимірювань. Тому вимірювання вважається закінченим тільки в тому випадку, якщо відомо, з якою похибкою воно здійснене.

2.1. Класифікація похибок вимірювань

Класифікацію похибок вимірювання проводять в залежності від джерела виникнення, умов проведення, характеру величини, що вимірюється в часі, та способу вираження.



За джерелом виникнення

Похибки методу вимірювання – спричинені недосконалістю цього методу, а також недостатністю обґрунтування його теорії, застосуванням наближених формул для спрощення розрахунків тощо.

Іструментальна похибка – складова похибки вимірювання – зумовлена недосконалістю засобів вимірювальної техніки (ЗВТ). Ця похибка також може бути обумовлена конструктивними та технологічними недоліками. Наприклад, через неточність виготовлення та нестабільності елементів ЗВТ, неправильне градування шкали приладу тощо.

Суб'єктивні (особисті) похибки – як правило, є наслідком особистих властивостей спостерігача (експериментатора), які зумовлені особливостями його організму (недосконалість зору, втомленість т

За умовами проведення вимірювань

Основна похибка – похибка, яка виникає за нормальних умов застосування ЗВТ ($20\pm 5^{\circ}\text{C}$, 750 ± 30 мм. АТ. ст., $65\pm 15\%$ - вологості). Ця похибка нормується і вказується у відповідних документах (технічному паспорті, формулярі).

Додаткова похибка – обумовлюється відхиленням однієї чи декількох впливових величин (температури, тиску, вологості тощо) від нормального значення. Значення додаткової похибки, як і основної, нормується і вказується у відповідних технічних документах.

За характером виявлення

Систематична похибка – складова похибки, яка залишається сталою або закономірно змінюється при повторних вимірюваннях однієї і тієї ж величини. Вона зумовлена впливом на результат вимірювання багатьох факторів, дію яких не усунуто та не прийнято до уваги. Ці фактори можуть бути або постійно діючими, або закономірно змінюватись. В ряді випадків систематичні похибки можуть бути визначені дослідним шляхом. В залежності від причин виникнення систематичні похибки поділяють на інструментальні, суб'єктивні, а також похибки методу та похибки від зовнішніх впливів. На практиці повне усунення систематичних похибок неможливе, отже, результат будь-якого вимірювання містить залишки не виключених систематичних похибок.

Випадкова похибка – це та складова похибки, яка за повторних вимірювань однієї й тієї ж величини, проведених за допомогою одного й того ж приладу, в однакових умовах, з однаковою старанністю, дасть результати спостережень, що мають (хоч незначно) відрізнятись один від одного. Це вказує на те, що при багаторазових вимірюваннях результати спостережень та їх похибки є випадковими величинами.

Поява випадкових похибок зумовлена спільним впливом на засіб та об'єкт вимірювання багатьох випадкових факторів, поміж якими відсутній взаємний зв'язок. Іншими словами, результат будь-якого вимірювання “обтяжений” випадковими похибками.

Груба похибка – це похибка вимірювання, яка істотно перевищує очікувану за даних умов похибку.

За характером зміни вимірюваної величини у часі

Статична похибка – похибка при вимірюванні постійної в часі величини. Наприклад, похибка вимірювання постійного струму тощо.

Динамічна похибка – різниця між похибкою в динамічному режимі (похибка при вимірюванні змінної в часі величини) і статичною похибкою, яка відповідає значенню вимірюваної величини у відповідний момент часу.

За способом виразу

Абсолютна похибка вимірювання — це алгебраїчна різниця між отриманим при вимірюванні значенням та істинним значенням вимірюваної величини:

$$\Delta x = x_{\text{в}} - X \quad (2.1)$$

де: Δx – абсолютна похибка вимірювання; $x_{\text{в}}$ – результат вимірювання;

X – істинне значення вимірюваної величини.

Оскільки істинне значення вимірюваної величини невідоме, то його замінюють на дійсне (яке має бути наближеним до істинного). Таким чином, Δx визначається:

$$\Delta x = x_{\text{в}} - X_{\text{д}} \quad (2.2)$$

де $X_{\text{д}}$ – дійсне значення вимірюваної величини.

Абсолютна похибка визначається в одиницях величини, яку вимірюють;

Відносна похибка – відношення абсолютної похибки вимірювання до істинного значення вимірюваної величини:

$$\delta_x = \frac{\Delta x}{X} \cdot 100\% \quad (2.3)$$

Відносна похибка виражена в безрозмірних одиницях (або у відсотках).

На практиці замість істинного значення використовують дійсне значення;

Приведена похибка – відношення абсолютної похибки до нормуючого значення вимірюваної величини:

$$\gamma = \frac{\Delta x}{X_{\text{нор}}} \quad (2.4)$$

де $X_{\text{нор}}$ - нормуюче значення.

Нормуюче значення приймають рівним:

для засобів вимірювань, у яких нульова відмітка знаходиться на краю або за межами шкали, – кінцевому значенню діапазону вимірювань;

якщо нульова відмітка знаходиться в межах діапазону вимірювань, – сумі кінцевих значень діапазону вимірювань;

для вимірювальних приладів з суттєвою нерівномірністю шкали нормуюче значення встановлюють рівним в усій довжині шкали або її частині, відповідній до діапазону вимірювань.



Рис. Загальна класифікація вимірювань.

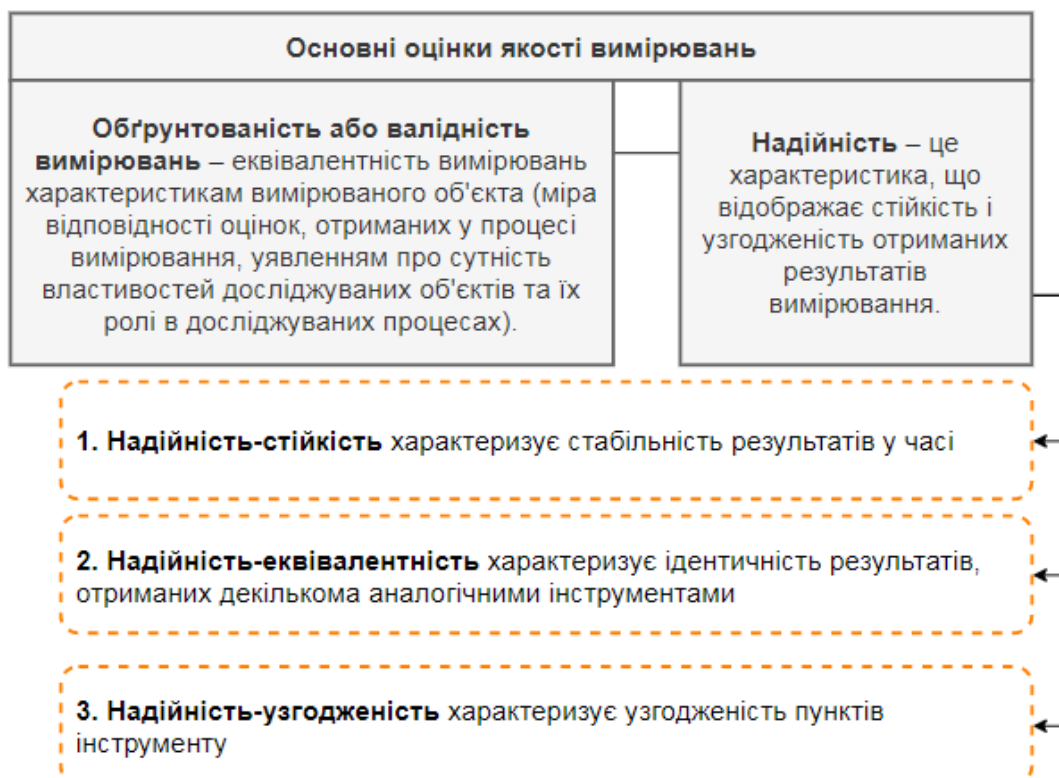


Рис. Основні оцінки якості вимірювань.



Рис. Типи валідності.

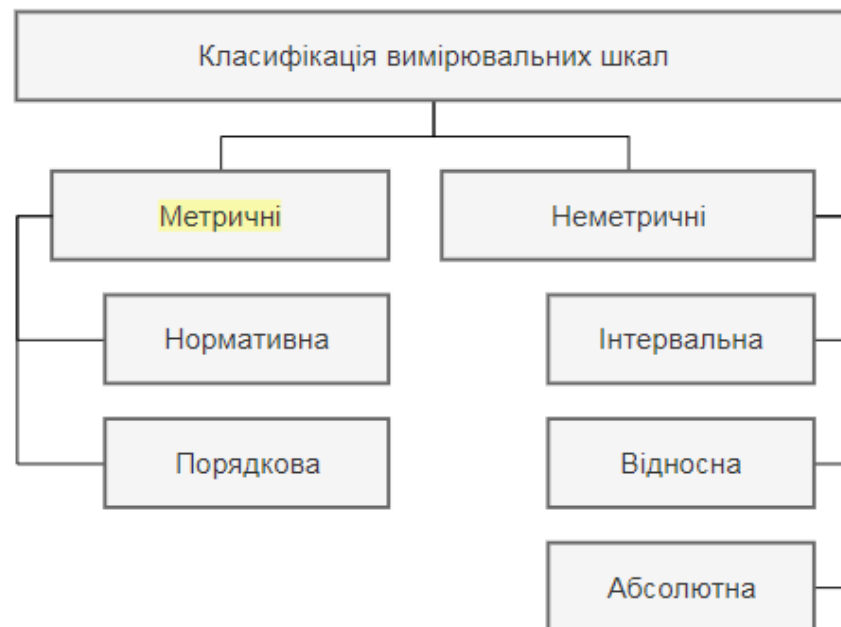


Рис. Класифікація вимірювальних шкал.

Теми доповідей для заслуховування:

1. Класифікація видів та методів вимірювання.
2. Метрологічні характеристики засобів вимірювання.
3. Основи теорії похибок вимірювань.
4. Міжнародна система одиниць фізичних величин.
5. Класифікація та характеристики похибок.
6. Розрахунок похибок при прямих вимірах.
7. Розрахунок похибок при непрямих вимірах.
8. Систематичні похибки та їх корекція.

Питання для контролю знань:

1. Що таке фізична величина?
2. На які значення поділяють фізичні величини?
3. Система одиниць вимірювання.
4. Міжнародна система одиниць.
5. Які існують вимірювання за характером знаходження значення величини, що вимірюють?
6. Класифікація методів вимірювання фізичних величин.
7. Дайте загальне визначення похибки вимірювань.
8. Які існують похибки за умов проведення вимірювань?
9. Які існують похибки за характером виявлення?
10. Дайте визначення похибкам від тимчасового характеру вимірюваної величини.

Завдання для самостійної підготовки:

- I. Вивчити матеріал попередньої лекції та практичного заняття.
- II. Пройти тестування: тест 2.2. за посиланням <https://forms.gle/uRLttRa7Pz3bqvWMA>

ПРАКТИЧНЕ ЗАНЯТТЯ № 3

КЛАСИФІКАЦІЯ ДАНИХ ТА СПОСОБИ ЇХ ОТРИМАННЯ

Література:

1. Карагодова О.О., Рожок В.Д. Дослідження операцій: навч. посіб. К.: Центр учбової літератури, 2007. 256 с.
2. Бахрушин В.Є. Методи аналізу даних: навчальний посібник для студентів. Запоріжжя: КПУ, 2011. 268 с
3. Перегуда О.В., Капустян О.А., Курилко О.Б. Статистична обробка даних: навч. посіб. Електронне видання, 2022. 103 с.
4. Рогальський Ф. Б., Курилович Я. Є., Цокурєнко А. А. Математичні методи аналізу економічних систем. Книга 1. Київ: Наукова думка, 2001, 435 с.
5. Рогальський Ф. Б., Цокурєнка А. А. Математичні методи аналізу економічних систем. Книга 2. Київ: Наукова думка, 2001, 423 с.
6. Крайчук О.В., Московська Г.К., Соколенко О.І. Теорія ймовірностей і математична статистика. Рівне, 2004.

Мета:

1. Навчальна – вивчити класифікацію даних та способи їх отримання для подальшої обробки та представлення.
2. Виховна – виховувати професійні риси, почуття відповідальності; виховувати зацікавленість дисципліною, прагнення отримати нові знання самостійно.
3. Розвиваюча – прививати навички наукового підходу у вирішенні задач забезпечення пожежної безпеки.

Питання для обговорення:

1. Типи вхідних даних.
2. Класифікація даних.
3. Завдання та методи класифікації даних.

Теоретична частина:

Дані - це сукупність фактів, пов'язаних із числами, символами, зображеннями, а також словами.

З технічної точки зору, класифікація даних — це процес організації даних за відповідними категоріями, такими як метадані, тип файлу та вміст. Це дозволяє користувачам знаходити свій набір даних більш ефективно та точно. Класифікація даних також впроваджується для захисту пов'язаних даних, оскільки різні форми даних можуть вимагати унікальних заходів захисту.

1. Типи вхідних даних.

Характеристичний опис – найпоширеніший випадок. Кожен об'єкт описується набором своїх характеристик, які називаються ознаками. Ознаки можуть бути числовими або нечисловими.

Матриця відстаней між об'єктами. Кожен об'єкт описується відстанями до всіх інших об'єктів навчальної вибірки. З цим типом вхідних даних працюють деякі методи, зокрема, метод найближчих сусідів, метод потенційних функцій.

Часовий ряд або сигнал є послідовність вимірів у часі. Кожен вимір може представлятися числом, вектором, а в загальному випадку – характеристичним описом досліджуваного об'єкта в цей час часу.

Зображення або відеоряд.

Зустрічаються і складніші випадки, коли вхідні дані представляються у вигляді графів, текстів, результатів запитів до бази даних, і т. д. Як правило, вони приводяться до першого або другого випадку шляхом попередньої обробки даних та вилучення характеристик.

2. Класифікація даних.

Класифікація даних на основі типів даних.

Обчислені дані – це отримані дані за результатами розрахунку. Що стосується того, що включено до обчислених даних, то це відсоток, пов'язаний з певним обсягом даних.

Дані вимірювання – це дані, які вказують на певний розмір або значення. Це значення може бути у вигляді цифр або букв, які, як правило, дають начальство для адресації працівникам після того, як керівник раніше проводив різні перевірки звітів про результати роботи цих працівників.

Класифікація даних виходячи з характеру даних.

Кількісні дані – це класифікація дані на основі суми.

Якісні дані – це дані протилежністю кількісних даних. Якісні дані – це дані, класифікація яких отримана на основі зв'язків чи відносин, що стосуються якостей, якими володіє хтось чи щось.

Класифікація даних на основі джерел даних

Внутрішні дані – це вихідні дані або справжні. Це внутрішні дані, отримані за результатами досліджень, проведених ним самим. Дослідження проводилось без допомоги чи втручання з боку інших. Тому ці дані є оригінальними або суто його власними роботами.

Зовнішні дані – це дані, яких немає справжні. Тут не оригінально, оскільки дані були отримані на основі оцінок, зроблених іншими. Отже, хтось може просто використовувати ці дані для своїх цілей, навіть якщо це в основному не самі дані, а дані, зібрані іншими людьми. Теоретично зовнішні дані діляться на два, включаючи:

Первинні зовнішні дані або зовнішній праймер даних. Ці дані є суміжними даними з усним мовленням у формі слів інших людей, які є власниками даних. Тобто ці дані отримують із інформації того, хто є оригінальним дослідником даних.

Вторинні зовнішні дані. Ці дані є отриманими даними від особи, яка не є власником даних. Отже, людина, яка використовується в якості посилання в цих даних, походить не від дослідників, а від інших людей, які також можуть отримати інформацію про ці дані від інших людей тощо.

Класифікація даних за часом збору

Дані переріз це дані, що показують інформацію щодо певного моменту.

Дані часовий ряд який також називається періодичними даними колекція, посилаючись на період або час від часу історично.

3. Завдання та методи класифікації даних.

У загальному випадку класифікацією (розпізнаванням образів) називають поділ досліджуваної сукупності об'єктів на однорідні в певному розумінні групи (класи) або зарахування кожного із заданої множини об'єктів до деякого із задалегідь відомих класів. При цьому вирізняють три групи завдань: дискримінацію, кластеризацію й групування. Останні дві групи є близькими за метою (поділ даних на класи або групи близьких у певному розумінні об'єктів), а також за алгоритмами. Але принципова різниця між ними полягає у тому, що у першому випадку межі класів є природними, а у другому – умовними й їх можна встановлювати суб'єктивно.

При побудові методів розв'язування цієї задачі зазвичай прагнуть мінімізувати ймовірність неправильної класифікації. Для цього можна побудувати функцію втрат $c(j|i)$, що характеризує втрати від помилкового зарахування об'єкта i -го класу до j -го класу [4]. При $i = j$ беруть $c(j|i) = 0$, а при $i \neq j$ – $c(j|i) > 0$. Якщо кількість таких помилок є $m(j|i)$, то загальні втрати:

$$C_n = \sum_{i=1}^k \sum_{j=1}^k c(j|i)m(j|i),$$

де n – кількість класифікованих об'єктів; k – кількість класів.

Величина C_n залежить від n . Для усунення такої залежності можна ввести питомі втрати (у розрахунку на один об'єкт) і перейти до границі при $n \rightarrow \infty$. Тоді:

$$C = \lim_{n \rightarrow \infty} (C_n / n) = \sum_{i=1}^k \pi_i \sum_{j=1}^k c(j|i)P(j|i) = \sum_{i=1}^k \pi_i C^{(i)},$$

де π_i – апіорна ймовірність (питома вага) i -го класу, $P(j|i)$ – ймовірність помилкового зарахування об'єкта i -го класу до j -го класу,

величина $C(i)$ визначає середні втрати від неправильної класифікації об'єктів i -го класу.

У багатьох випадках втрати є однаковими для будь-якої пари i та j , тобто:

$$c(j|i) = c_0 = \text{const} \quad (i \neq j).$$

Тоді мінімізація функції втрат еквівалентна максимізації ймовірності правильної класифікації, яка дорівнює $\sum_{i=1}^k \pi_i P(i|i)$. З огляду на це, при побудові процедур класифікації часто розв'язують задачу мінімізації ймовірності неправильної класифікації:

$$\frac{C}{c_0} = 1 - \sum_{i=1}^k \pi_i P(i|i).$$

На результат класифікації істотно впливають типи класів. Найчастіше виокремлюють такі типи:

1. **Клас типу ядра, або згущення.** У цьому випадку всі відстані між об'єктами всередині класу є меншими, ніж їх відстані до будь-якого об'єкта, що не входить до цього класу.

2. **Кластер, або згущення у середньому.** Середня відстань між об'єктами всередині класу є меншою, ніж їх середня відстань до всіх інших об'єктів.

3. **Для класу типу стрічки** існує таке $\varepsilon > 0$, що в цьому класі є хоча б один об'єкт x_i , для якого відстань до будь-якого іншого об'єкта цього класу x_j $d_{ij} < \varepsilon$, а відстань до будь-якого об'єкта x_k , що не належить до цього класу, $d_{ij} < \varepsilon$.

4. Характеристичною властивістю **класу із центром** є існування певних граничного значення R і точки x^* у просторі ознак таких, що у багатовимірному шарі радіуса R з центром в точці x^* містяться всі елементи цього класу й немає елементів, які не належать до нього.

Слід зазначити, що один й той самий клас може задовольняти визначення декількох типів. Тому вказана класифікація важлива не стільки з погляду зарахування конкретного класу до певного типу, а з погляду вибору методів розділення декількох класів. Класи, що перетинаються, можуть не задовольняти будь-які з наведених вимог. Але в окремих випадках їх також можна розділяти за допомогою формальних алгоритмів класифікації.

Теми доповідей для заслуховування:

1. Параметричні методи класифікації без навчання.
2. Сутність ієрархічних методів кластерного аналізу.
3. Методи розпізнавання образів з навчанням.
4. Методи розпізнавання образів без навчанням.
5. Базовий алгоритм методу середнього зв'язку Кінга.

6. Базовий алгоритм методу k -середніх Мак-Куїна.
7. Сутність дискримінантного аналізу.

Питання для контролю знань:

1. Надайте визначення термінам: «дані», «класифікація даних».
2. Що називають класифікацією даних?
3. Який критерій найчастіше застосовують для розробки процедур класифікації даних?
4. Класифікація даних на основі типів даних.
5. Класифікація даних на основі джерел даних.
6. Класифікація даних за часом збору.
7. Завдання класифікації даних.
8. Методи класифікації даних.

Завдання для самостійної підготовки:

- I. Вивчити матеріал лекції.
- II. Підготувати доповідь за темою:
 1. Форми звітності при науковому дослідженні.
 2. Наукова публікація: поняття, функції, основні види.
 3. Наукова монографія, наукова стаття, тези.
 4. Реферат, доповідь, виступ.
 5. Методика підготовки та оформлення публікацій.
- III. Пройти тестування за темою лекції: тест 3.2. за посиланням <https://forms.gle/1vFHJDBHjY9h62cKA>

ПРАКТИЧНЕ ЗАНЯТТЯ № 4

ЗВІТНІСТЬ ТА ПУБЛІКАЦІЯ РЕЗУЛЬТАТІВ ЕКСПЕРИМЕНТУ

Література:

1. Карагодова О.О., Рожок В.Д. Дослідження операцій: навч. посіб. К.: Центр учбової літератури, 2007. 256 с.
2. Бахрушин В.Є. Методи аналізу даних: навчальний посібник для студентів. Запоріжжя: КПУ, 2011. 268 с.
3. Перегуда О.В., Капустян О.А., Курилко О.Б. Статистична обробка даних: навч. посіб. Електронне видання, 2022. 103 с.
4. Рогальський Ф. Б., Курилович Я. Є., Цокурєнка А. А. Математичні методи аналізу економічних систем. Книга 1. Київ: Наукова думка, 2001, 435 с.
5. Рогальський Ф. Б., Цокурєнка А. А. Математичні методи аналізу економічних систем. Книга 2. Київ: Наукова думка, 2001, 423 с.
6. Крайчук О.В., Московська Г.К., Соколенко О.І. Теорія ймовірностей і математична статистика. Рівне, 2004.

Мета:

1. Навчальна – вивчити загальні відомості про методи обробки та експериментальних даних та їх представлення.
2. Виховна – виховувати професійні риси, почуття відповідальності; виховувати зацікавленість дисципліною, прагнення отримати нові знання самостійно.
3. Розвиваюча – прививати навички наукового підходу у вирішенні задач забезпечення пожежної безпеки.

Питання для обговорення:

Форми представлення результатів дослідження.

Теоретична частина:

Результати тієї чи іншої наукової діяльності дослідника тільки тоді мають практичну цінність і значущість, коли вони доведені до відому членів суспільства. Однією із найпоширеніших форм такої діяльності є наукові публікації, які виконують наступні функції:

- оприлюднюють результати наукової роботи автора і сприяють встановленню його пріоритету;
- свідчать про відповідний особистий внесок дослідника в розробку наукової проблеми;
- підтверджують достовірність основних результатів і висновків, новизни і наукового рівня досліджень;
- підтверджують факт апробації та впровадження результатів і висновків наукової роботи;

- відбивають основний зміст наукової роботи;
- фіксують завершення певного етапу дослідження або роботи в цілому;
- забезпечують первинною науковою інформацією суспільство, сповіщають наукове співтовариство про появу нового наукового знання, передають індивідуальний результат у загальне надбання тощо.

Завершенням будь-якої дослідницької роботи є представлення результатів у тій формі, яка прийнята науковим співтовариством. Слід розрізняти дві основні форми представлення результатів: кваліфікаційну та науково-дослідницьку.

Кваліфікаційна робота – курсова, дипломна робота, дисертація тощо. Служить для того, щоб здобувач вищої освіти, представивши своє наукове дослідження, отримав документ, що засвідчує рівень компетентності. Вимоги до таких робіт, способу їх оформлення та подання результатів викладені у відповідних інструкціях і положеннях, прийнятих вченими радами.

Результати **науково-дослідної роботи** – це результати, отримані в ході дослідницької діяльності вченого. Подання наукових результатів зазвичай відбувається в трьох формах: 1) усні викладу; 2) публікації; 3) електронні версії. У будь-який з цих форм є опис. Під описом розуміємо будь-яку форму подання інформації про отримані в дослідженні результати.

Розрізняють такі варіанти представлення інформації: вербальна форма (текст, мова), символічна (знаки, формули), графічна (схеми, графіки), предметнообразна (макети, речові моделі, фільми та ін.).

Вербальна форма – найбільш поширений варіант подання описів. Будь-яке наукове повідомлення – це перш за все текст, організований за певними правилами. Розрізняють два види текстів: на природному мовою («природному», повсякденному) і на науковому мовою. Зазвичай представлення результатів наукового дослідження є текстом «змішаного» виду, де в природну мовну структуру включені фрагменти, сформульовані на строго науковому мовою. Ці мови не можна строго розмежувати: наукові терміни входять в повсякденне звернення, а наука черпає з природної мови слова для позначення знову відкритих сторін реальності. Але на відміну від звичайного вживання кожен науковий термін має однозначне предметний зміст.

Головна вимога до наукового тексту – послідовність і логічність викладу. Автор повинен по можливості не завантажувати текст надлишковою інформацією, але може використовувати метафори, приклади, для того щоб привернути увагу до особливо значущого для розуміння суті ланці міркувань. Науковий текст на відміну від літературного тексту або повсякденної мови дуже клішірований - в ньому переважають стійкі структури і обороти (в цьому він схожий з «канцеляритом» - бюрократичною мовою ділових паперів). Роль таких

штампів надзвичайно важлива, оскільки увагу читача не відволікається на літературні шедеври або неправильності викладу, а зосереджується на значимій інформації: судженнях, умовиводах, доказах, цифрах, формулах. «Наукоподібні» штампи насправді відіграють важливу роль «рамки», стандартної установки для нового наукового змісту.

Текст складається з висловлювань. Кожне висловлювання має певну логічну форму. Існують основні логічні форми висловлювання: 1) індуктивний – узагальнююче деякий емпіричний матеріал; 2) дедуктивний – логічний висновок від загального до приватного або опис алгоритму; 3) аналогія – «трансдукція»; 4) тлумачення чи коментар – «переклад», розкриття змісту одного тексту за допомогою створення іншого.

Геометричні (просторово-образні) опису є традиційним способом кодування наукової інформації. Оскільки геометричний опис доповнює і пояснює текст, воно «прив'язане» до опису мовного. Геометричний опис наочно. Воно дозволяє одночасно представити систему відносин між окремими змінними, досліджуваними в експерименті. Інформаційна ємність геометричного опису дуже велика.

Для первинного представлення даних використовуються наступні графічні форми: діаграми, гістограми та полігони розподілу, а також різні графіки.

Початковим способом представлення даних є зображення розподілу. Для цього використовують гістограми та полігони розподілу. Часто для наочності розподіл показника в експериментальній і контрольній групах зображають на одному малюнку.

Гістограма – це «столбчатая» діаграма частотного розподілу ознаки на вибірці. При побудові гістограм на осі абсцис відкладають значення вимірюваної величини, а на осі ординат – частоти або відносні частоти народження даного діапазону величини у вибірці.

В полігоні розподілу кількість випробовуваних, що мають дану величину ознаки (або потрапили в певний інтервал величини), позначають крапкою з координатами. Точки з'єднуються відрізками прямої. Перед тим як будувати полігон розподілу або гістограму, дослідник повинен розбити діапазон вимірюваної величини, якщо ознака дан в шкалі інтервалів або відносин, на рівні відрізки. Рекомендують використовувати не менше п'яти, але не більше десяти градацій. У разі використання шкали найменувань або порядкової шкали такої проблеми не виникає.

Якщо дослідник хоче наочніше уявити співвідношення між різними величинами, наприклад частки випробовуваних з різними якісними особливостями, то йому вигідніше використовувати діаграму. У секторній круговій діаграмі величина кожного сектора пропорційна величині зустрічальності кожного типу. Величина кругової діаграми може відображати відносний обсяг вибірки або значимість ознаки.

Перехідним від графічного до аналітичного варіантом відображення інформації є в першу чергу графіки, що представляють функціональну

залежність ознак. Ідеальний варіант завершення експериментального дослідження – виявлення функціонального зв'язку незалежної і залежної змінних, яку можна описати аналітично.

Можна виділити два різних за змістом типу графіків: 1) відображають залежність зміни параметрів у часі; 2) відображають зв'язок незалежної та залежної змінних (або будь-яких двох інших змінних).

Рекомендації з побудови графіків:

1. Графік і текст повинні взаємно доповнювати один одного.
2. Графік повинен бути зрозумілий «сам по собі» і включати всі необхідні позначення.
3. На одному графіку не дозволяється зображати більше чотирьох кривих.
4. Лінії на графіку повинні відображати значимість параметра, найважливіші параметри необхідно позначати цифрами.
5. Написи на осях слід розташовувати внизу і зліва.
6. Точки на різних лініях прийнято позначати кружками, квадратами і трикутниками.

Якщо необхідно на тому ж графіку уявити величину розкиду даних, то їх слід зображати у вигляді вертикальних відрізків, щоб точка, що позначає середнє, перебувала на відрізку (відповідно до показника асиметрії).

Видом графіків є діагностичні профілі, які характеризують середню вираженість вимірюваних показників у групи або певного індивіда.

При поданні інформації з використанням топологічних характеристик застосовуються графи.

Найбільш важливий спосіб представлення результатів наукової роботи - числові значення величини, зокрема:

- 1) показники центральної тенденції (середнє, мода, медіана);
- 2) абсолютні та відносні частоти;
- 3) показники розкиду (стандартне відхилення, дисперсія);
- 4) значення критеріїв, використаних при порівнянні результатів різних груп;
- 5) коефіцієнти лінійної і нелінійної зв'язку змінних і т.д.

Стандартний вид таблиць для представлення первинних результатів такий: по рядках розташовуються випробовувані, по стовпцях - значення вимірюваних параметрів. Результати математичної статистичної обробки також зводяться в таблиці. Існуючі комп'ютерні пакети статистичної обробки даних дозволяють вибрати будь-яку стандартну форму таблиць для представлення їх в науковій публікації.

Теми доповідей для заслуховування:

1. Форми звітності при науковому дослідженні.
2. Наукова публікація: поняття, функції, основні види.
3. Наукова монографія, наукова стаття, тези.

4. Реферат, доповідь, виступ.
5. Методика підготовки та оформлення публікацій.

Питання для контролю знань:

1. Надайте визначення термінам «аналіз даних», «спостереження», «експеримент», «імітація», «попередня обробка даних».
2. Основні форми представлення результатів.
3. Функції наукової публікації.
4. Варіанти представлення інформації.
5. Спосіб представлення результатів наукової роботи.
6. Методи статистичної обробки результатів експерименту.
7. Первинні методи математико-статистичного аналізу.
8. Вторинні методи математико-статистичного аналізу.
9. Значення понять: «мода», «медіана», «вибіркове середнє», «розкид вибірки», «дисперсія».

Завдання для самостійної підготовки:

- I. Вивчити матеріал попередньої лекції.
- II. Пройти тестування: тест 4.2. за посиланням <https://forms.gle/NotAL5sCV1FrnRZj8>

ПРАКТИЧНЕ ЗАНЯТТЯ № 5

ПРЕДСТАВЛЕННЯ РЕЗУЛЬТАТІВ ОБРОБКИ ДАНИХ

Література:

1. Карагодова О.О., Рожок В.Д. Дослідження операцій: навч. посіб. К.: Центр учбової літератури, 2007. 256 с.
2. Бахрушин В.Є. Методи аналізу даних: навчальний посібник для студентів. Запоріжжя: КПУ, 2011. 268 с
3. Перегуда О.В., Капустян О.А., Курилко О.Б. Статистична обробка даних: навч. посіб. Електронне видання, 2022. 103 с.
4. Рогальський Ф.Б., Курилович Я.Є., Цокурєнко А.А. Математичні методи аналізу економічних систем. Книга 1. Київ: Наукова думка, 2001, 435 с.
5. Рогальський Ф.Б., Цокурєнка А.А. Математичні методи аналізу економічних систем. Книга 2. Київ: Наукова думка, 2001, 423 с.
6. Крайчук О.В., Московська Г.К., Соколенко О.І. Теорія ймовірностей і математична статистика. Рівне, 2004.

Мета:

1. Навчальна – вивчити загальні відомості про методи обробки та експериментальних даних та їх представлення.
2. Виховна – ознайомитися з методами обробки експериментальних даних та їх представлення.
3. Розвиваюча – прививати навички наукового підходу у вирішенні питань застосування методів обробки експериментальних даних та їх представлення.

Питання для обговорення:

1. Попередня обробка даних.
2. Представлення результатів обробки даних.

Теоретична частина:

1. Попередня обробка даних.

Попередня обробка даних – це техніка обміну даними, яка передбачає перетворення необроблених даних у зрозумілий формат. Дані в реальному світі часто неповні, непослідовні та/або відсутні у певній поведінці чи тенденціях і, ймовірно, містять багато помилок.

Попередня обробка даних – це перевірений метод вирішення таких питань. Попередня обробка даних готує вихідні дані для подальшої обробки.

Попередня обробка даних використовується додатками, керованими базами даних, такими як управління відносинами з клієнтами та додатки на основі правил (наприклад, нейронні мережі).

Під час попередньої обробки дані проходять ряд кроків:

- Очищення даних: дані очищаються за допомогою таких процесів, як заповнення пропущених значень, згладжування галасливих даних або усунення невідповідностей у даних.
 - Інтеграція даних: дані з різними поданнями складаються разом і конфлікти всередині даних вирішуються.
 - Трансформація даних: дані нормалізуються, агрегуються та узагальнюються.
 - Скорочення даних: цей крок спрямований на подання зменшеного представлення даних у сховищі даних.
 - Дискретизація даних: включає в себе зменшення ряду значень безперервного атрибута шляхом ділення діапазону інтервалів атрибутів.
- Завдання попередньої обробки даних:
- очищення даних;
 - редагування даних;
 - зменшення даних.

2. Форми представлення результатів дослідження.

Для запису результатів великої кількості однотипних вимірів використовують таблиці. З їхньою допомогою вдається уникнути непотрібного багаторазового запису позначення вимірюваної величини, одиниць виміру, використовуваних масштабних коефіцієнтів і т.п. У таблиці, крім експериментальних даних, зводяться проміжні результати обробки цих даних. У заголовок таблиці заносяться розмірності величин, характерні ступені. У таблиці вказується порядковий номер кожного виміру.

Представлені у табличній формі результати експериментів не завжди дозволяють наочно характеризувати закономірності досліджуваних процесів. Тому, при обробці результатів вимірювальної інформації, часто використовують методи графічного зображення. Графічне зображення дає найбільш наочне подання результатів експерименту, дозволяє краще зрозуміти фізичну сутність досліджуваного процесу, виявити загальний характер функціональної залежності досліджуваних змінних величин, установити наявність екстремумів. Для графічного зображення результатів вимірів найчастіше застосовують систему прямокутних координат. Осі графіка у прямокутній системі координат: по горизонтальній осі (осі абсцис) відкладають аргумент, незалежну величину, а по вертикальній осі (осі ординат) – функцію, залежну величину. Масштаб по осях – чисельне значення фізичної величини, що відповідає одиничному відрізку. Осі обов'язково повинні містити початок координат – звичайно враховують мінімальне й максимальне значення.

Шкала – підписи до осей у вигляді числового масштабу, з урахуванням масштабного коефіцієнта. Звичайно вибираються якісь “круглі”, зручні числа з мінімумом знаків після коми.

Масштабна сітка – для зручності визначення величин конкретних точок роблять тонкі вертикальні й горизонтальні лінії, які є продовженнями оцінок шкали.

Погрішності вимірів – навколо проставленої експериментальної точки будують два відрізки δy і δx , паралельні осям абсцис й ординат. В обраному масштабі довжина кожного відрізка повинна рівнятися подвоєної погрішності величини, що відкладає по паралельній осі. Центр відрізка повинен доводитися на експериментальну точку.

Назва – під графіком повинна бути наведена його назва, яка пояснює, до чого ставиться зображена залежність.

Всі сторінки, таблиці, формули, схеми й графіки повинні бути пронумеровані (у порядку використання). На початку звіту звичайно приводять зміст звіту. Якщо таблиці або графіки мають значний розмір і заважають зв'язаному сприйняттю тексту, їх варто винести в “Додатки” і дати на них посилання в тексті.

У деяких випадках можливе використання косокутних систем координат. Такий прийом дозволяє наочно зобразити досліджуваний процес і відбити взаємозв'язок декількох параметрів.

Якщо графічним методом аналізується функція $y = f(x)$, то в системі прямокутних координат наносяться точки (рис. 3.2, а) з отриманими координатами $\{x_0, y_0\}$, $\{x_1, y_1\}$, ..., $\{x_n, y_n\}$.

Перед побудовою графіка необхідно з'ясувати вплив досліджуваного явища, базуючись на відомих теоретичних дослідженнях або на попередній оцінці експериментальних даних.

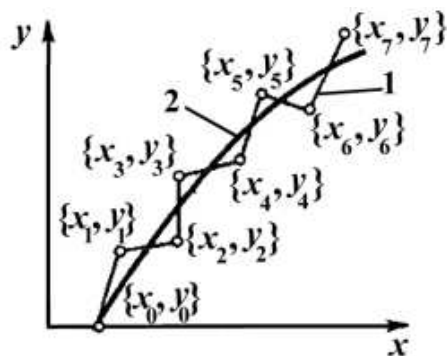


Рис. 3.2, а. Графічне зображення залежності $y = f(x)$:
1 – крива за результатами безпосередніх вимірів;
2 – фактична крива

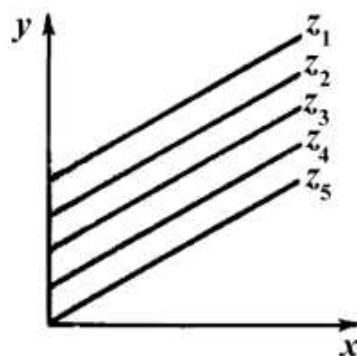


Рис. 3.2, б. Графічне зображення сімейства кривих при наявності декількох змінних

При вимірах фізичних величин часто зручно проводити виміри графічним методом, визначаючи безпосередньо із графіків деякі параметри. Для лінійних залежностей із графіка визначають тангенс кута нахилу й координати точок перетинання з осями. Графіки зручно будувати на аркушах міліметрового паперу. По обох осях масштаби графіку

потрібно вибирати такими, щоб передбачувані залежності мали найбільшу наочність і заповнювали більшу частину поля. Як правило, при використанні рівномірних шкал, сітка графіка повинна бути квадратною.

Рекомендується використати масштаби 1:1; 1:10; 1:100 і т.д., 1:2; 1:20; 1:200 і т.д., 1:5; 1:50; 1:500 і т.д. Ці масштаби найбільш зручні для перерахування вимірюваних величин в одиниці довжини, що відкладають по осях. Стрілки на кінцях осей графіка, як правило, не ставлять, але обов'язково вказують позначення фізичних величин і одиниці їхнього виміру. Якщо значення фізичної величини містять множники 10^n , то їх відносять до одиниці виміру. З огляду на те, що міліметровий папір має дуже дрібну сітку, на графік варто нанести тільки велику сітку. Оцифровують лише розподіли великої сітки.

Для нанесення отриманих точок використовують умовні позначки: світлі й темні кружки, квадратики, трикутники, хрестики й т.п. При побудові графіків варто мати на увазі, що за результатами експериментів одержуємо не точку, а прямокутник зі сторонами δy і δx , тому необхідно проводити плавну лінію так, щоб приблизно однакове число точок виявилось, по різні сторони від кривої.

Якщо передбачувана залежність – лінійна, то зручно проводити лінію за допомогою прозорої лінійки. Однак, найкращі результати виходять при застосуванні методу найменших квадратів (МНК), відповідно до якого сума квадратів відхилень експериментальних точок, від точок лежачих на кривій, повинна бути мінімальною. Сутність МНК укладена у тому, що відшукуються такі значення x , при яких мінімізується сума квадратів відхилень (помилки):

$$e_i = y_i - f(x_i): \sum_{i=1}^N e_i^2 = \sum_{i=1}^N (y_i - f(x_i))^2 \rightarrow \min_x$$

Якщо точки на графіку з'єднати прямими відрізками, то одержимо ламану криву, що характеризує зміну функції безпосередньо за даними експерименту. Однак функції можуть мати плавний характер, і якби експеримент повторити з використанням засобів вимірів більш високої точності й більшою кількістю точок виміру, то ламана крива більше б відповідала плавній кривій. Тому при побудові функції, лінія що її відображає повинна проходити по можливості ближче до всіх експериментальних точок.

Якщо досліджуються явища, для яких у певних інтервалах зміни x спостерігається швидка або стрибкоподібна зміна однієї з координат необхідно особливо ретельно підійти до побудови кривої, що б уникнути підміни стрибка функції погрішностями вимірів.

Про наявність “стрибка” функції може інформувати різке віддалення від кривій декількох послідовних точок. У такому випадку необхідно проаналізувати фізичну сутність явища. Якщо немає підстави для припущення про наявність “стрибка” функції, то таке різке відхилення можна пояснити грубою помилкою або промахом виміру. Варто повторити

вимір у діапазоні різкого відхилення даних. Якщо повторні виміри дадуть результати близькі до первісних, то робиться вивід про наявність “стрибка” на розглянутому інтервалі кривої, який повинен бути пояснений фізичною сутністю явища.

При графічному відображенні результатів експерименту часто необхідно розглядати три змінних $\beta = f(x, y, z)$. У такому випадку звичайно застосовують метод поділу змінних. Сутність методу полягає у тому, що однієї з величин z у межах інтервалу вимірів z_1, \dots, z_n задають кілька фіксованих значень. Для двох останніх змінних x та y будують графіки $y = f(x)$ при $z_i = \text{const}$. У результаті, на одному графіку одержують сімейство кривих $y = f(x)$ для різних значень z .

Іноді необхідно графічно зобразити функцію із чотирма й більше змінними, наприклад, $\alpha = f(x, y, z, b)$. У цьому випадку будується серія графіків, але кожний із графіків при будують при деякому значенні $b_i = \text{const}$, або при n -змінних приймається постійними $(n - 1)$ змінних й будуються графіки $(n - 1) = f_1(x)$, далі $(n - 2) = f_2(x)$, $(n - 3) = f_3(x)$, й т.д., що дозволяє простежити зміну функції від іншої змінної, але при постійних значеннях інших змінних. Такий метод графічного аналізу хоча й вимагає великого обсягу вимірів, старанності й уваги до отриманих результатів, проте, у більшості випадків, є найбільш простим і наочним. Існуючі сучасні пакети математичної обробки та моделювання, наприклад – Matlab, дозволяють суттєво автоматизувати процеси як отримання експериментальних даних, так і їх графічного представлення, обробки та подальшого аналізу.

При графічному поданні результатів експериментів велике значення має вибір координатної сітки. Координатні сітки можуть бути рівномірними й нерівномірними. У рівномірних координатних сіток ординати й абсциси мають рівномірні шкали, тобто довжини одиничних відрізків, що відкладають на осі, однакові.

З нерівномірних координатних сіток найпоширеніші напівлогарифмічна, логарифмічна й імовірнісна. У більшості випадків нерівномірні координатні сітки використовують для більш наочного зображення функцій. Графік функції $y = f(x)$ при його відображенні в різних сітках має різну форму. При використанні логарифмічної сітки багато криволінійних функцій спрямляються.

Напівлогарифмічна сітка має логарифмічну ординату й рівномірну абсцису або рівномірну ординату й логарифмічну абсцису. Логарифмічна сітка має обидві логарифмічні осі. Імовірнісна координатна сітка, звичайно, має рівномірну ординату й імовірнісну шкалу абсцис.

Особливе місце в практиці графічного зображення експериментальних даних займає імовірнісна сітка. Застосування такої сітки стає доцільним у ході обробці вимірів, наприклад, для різноманітних оцінок точності, для визначення розрахункових характеристик вологості, для визначення міжремонтних періодів і т.д. Побудова функції в

імовірнісній системі координат дозволяє виділити ті її ділянки, які є більше важливими при аналізі характеру зміни досліджуваної величини (параметра).

При обробці експериментальних даних іноді графічним способом необхідно побудувати розрахункові графіки з метою спрощення процесу знаходження по одній відомій змінній інших змінних. При цьому підвищуються вимоги до точності графічної побудови функції.

При побудові таких графіків необхідно:

- залежно від числа змінних вибрати координатні сітки;
- визначити вид графіка (одна крива, сімейство кривих, серія сімейств кривих);
- обрати такий масштаб графіку, що визначить його ефективне використання;
- для розрахункових графіків складного виду (маючих мінімуми або максимуми), кількість точок на ділянках вигину повинна бути значно більшою, ніж на плавних ділянках.

У деяких випадках будуються номограми, що полегшує виконання систематичних розрахунків складних теоретичних або емпіричних формул.

Номограми можуть відбивати алгебраїчні вираження й дозволяють вирішувати складні математичні вирази порівняно просто – графічними методами.

Існують різні методи побудови номограм із застосуванням рівномірних і нерівномірних координатних сіток.

Методика побудови номограм функції зводиться до побудови кривих, або їхнього сімейства, шляхом прийняття постійними окремих змінних. Складні алгебраїчні вираження доцільно зводити до простого добутку декількох значень, наприклад, $a \cdot b \cdot c$ де a , b й c – функції трьох змінних.

Спочатку необхідно, задавшись змінними, обчислити a , b й c . Далі, надаючи їм постійні значення визначити, наприклад, b . Величини a , b й c необхідно варіювати в певних межах через заданий інтервал Цей спосіб найбільш ефективний у випадку якщо a , b й c представляються, як безрозмірні критерії.

Теми доповідей для заслуховування:

1. Статистичні методи оцінки експериментальних даних.
2. Послідовність статистичного дослідження.
3. Статистична оцінка імовірності досліджень. Оцінка похибок вимірювання.
4. Статистичні критерії розходження.
5. Критерій Н.В. Смірнова.
6. Критерій Діксона.

Питання для контролю знань:

1. Надайте визначення термінам «аналіз даних», «спостереження», «експеримент», «імітація», «попередня обробка даних», «графік», «номограма».
2. Зміст попередньої обробки даних.
3. Завдання попередньої обробки даних.
4. Методи обробки експериментальних даних.
5. Основні форми представлення результатів.
6. Варіанти представлення інформації.
7. Спосіб представлення результатів наукової роботи.
8. Методи статистичної обробки результатів експерименту.
9. Первинні методи математико-статистичного аналізу.
10. Вторинні методи математико-статистичного аналізу.
11. Значення понять: «мода», «медіана», «вибіркове середнє», «розкид вибірки», «дисперсія».

Завдання для самостійної підготовки:

- I. Вивчити матеріал лекції.
- II. Підготувати доповідь за темою:
 1. Аналіз сучасного стану на ринку статистичних пакетів.
 2. Суть та організація статистичного зведення.
 3. Візуалізація та графічне відображення даних.
- III. Пройти тестування: тест 5.2. за посиланням <https://forms.gle/5kVKUmbq6jvRtRRz9>

ПРАКТИЧНЕ ЗАНЯТТЯ № 6 ПРЕДСТАВЛЕННЯ РЕЗУЛЬТАТІВ ОБРОБКИ ДАНИХ

Література:

1. Карагодова О.О., Рожок В.Д. Дослідження операцій: навч. посіб. К.: Центр учбової літератури, 2007. 256 с.
2. Бахрушин В.Є. Методи аналізу даних: навчальний посібник для студентів. Запоріжжя: КПУ, 2011. 268 с.
3. Перегуда О.В., Капустян О.А., Курилко О.Б. Статистична обробка даних: навч. посіб. Електронне видання, 2022. 103 с.
4. Рогальський Ф.Б., Курилович Я.Є., Цокурєнко А.А. Математичні методи аналізу економічних систем. Книга 1. Київ: Наукова думка, 2001, 435 с.
5. Рогальський Ф.Б., Цокурєнка А.А. Математичні методи аналізу економічних систем. Книга 2. Київ: Наукова думка, 2001, 423 с.
6. Крайчук О.В., Московська Г.К., Соколенко О.І. Теорія ймовірностей і математична статистика. Рівне, 2004.

Мета:

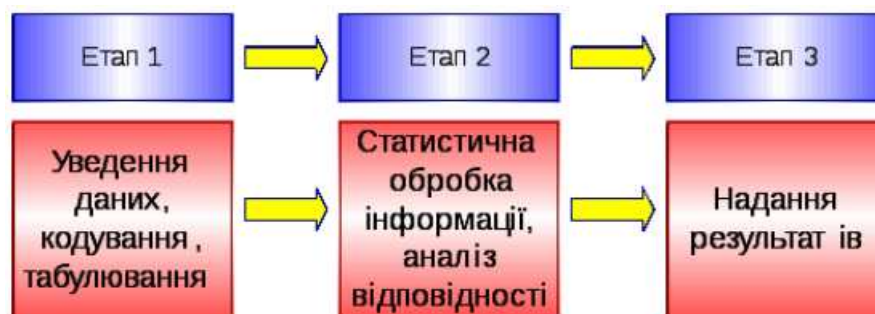
1. Навчальна – вивчити загальні відомості про методи обробки та експериментальних даних та їх представлення.
2. Виховна – виховувати професійні риси, почуття відповідальності; виховувати зацікавленість дисципліною, прагнення отримати нові знання самостійно.
3. Розвиваюча – прививати навички наукового підходу під час вирішення задач у сфері забезпечення пожежної безпеки.

Питання для обговорення:

Аналіз даних і представлення результатів дослідження.

Теоретична частина:

Аналіз даних – це процес послідовних логічних дій з інтерпретації відповідей респондентів в статистичні форми та інформацію, необхідну для ухвалення рішень. Зазвичай цей процес складається з трьох послідовних етапів (рис. 1).



На **першому** етапі дані, зібрані в процесі опитування вводяться в комп'ютер у вигляді матриць (відбувається табуляція даних). Уведення даних і їх табуляція може здійснюватися за допомогою спеціальних комп'ютерних програм, наприклад Vortex, SPSS, Statistica, або в середовищі MS Excel чи Access. І в одному і в другому випадку з'являється можливість сортування, форматування, відбору та інших операцій щодо обробки даних. Останнім часом з'явилися спеціальні програми обробки даних досліджень, в середовищі яких передбачене сканування анкет і подальша обробка отриманої інформації.

Слід зауважити, що введення даних є процесом трудомістким і часто вимагає роботи спеціально підготовлених фахівців. Тому ще на етапі планування дослідження доцільно визначитися з оптимальною (а краще – мінімальною) кількістю питань у анкеті, з метою запобігання труднощів під час введення та аналізу отриманих даних. Проблеми загострюються, коли опрацьовувати доводиться значні кількості анкет.

На **другому** етапі проводиться статистичний аналіз даних, пошук взаємозв'язків і відмінностей в масивах. Крім того, статистична обробка припускає розробку висновків і гіпотез (концептуалізація даних), одночасно проводиться перевірка репрезентативності результатів, їх здатності бути віднесеними на всю генеральну сукупність. Сам статистичний аналіз може проводитися по-різному (табл. 1).

Таблиця 1. Види статистичного аналізу даних

Вид аналізу	Характеристика
Дескриптивний	Оцінка результатів за допомогою узагальнюючих величин. Середні, мода, середньоквадратичне відхилення тощо.
Висновків	Використання статистичних методів з метою узагальнення отриманих результатів на всю сукупність.
Відмінностей	Використовується для оцінки відмінностей між двома групами даних (різних сегментів тощо)
Зв'язків	Пошук систематичних зв'язків між різними групами даних, наприклад, як ціна впливає на обсяг продажів тощо.
Прогнозний	Прогноз розвитку подій у майбутньому, наприклад, шляхом аналізу часових рядів.

Для проведення **дескриптивного аналізу** широко використовуються дві групи вимірювань.

Перша включає вимірювання «центральної тенденції» або опис типового респондента. Використовується:

- мода – величина ознаки, що з'являється найбільш часто;

- медіана – значення ознаки, що знаходиться в середині впорядкованого ряду;
- середнє (як правило, середнє арифметичне) – значення ознаки, що рівномірно розподілене між всіма одиницями сукупності.

Друга включає вимірювання варіації. Зстосовують:

- розподіл частот – показує кількість випадків появи кожного значення ознаки, яка вимірюється;
- розмах варіації – це різниця між кінцевими точками впорядкованого ряду значень;
- середнє квадратичне відхилення – узагальнююча характеристика варіації. Означає відсоток схожості або відмінностей оцінок.

Висновковий аналіз – вид аналізу, що спрямований на загальні припущення щодо сукупності даних на основі вивчення групи одиниць сукупності. Він визначає можливість розповсюдити висновки дослідження на всю сукупність і розмір похибки. Оцінка робиться на основі оцінки параметрів.

Оцінка параметрів – це визначення інтервалу в якому знаходиться один з параметрів сукупності.

Розрахунок параметрів у натуральних одиницях можна здійснити за формулами:

$$S_x = \frac{S}{\sqrt{n}}, \text{ де } S_x - \text{середня квадратична похибка вибіркового середнього};$$

S – середнє квадратичне відхилення від середнього у вибірці;

n – обсяг вибірки.

$$\bar{X} = X \pm ZS_x, \text{ де } \bar{X} - \text{середнє для генеральної сукупності}$$

X – вимірюваний параметр

Z – параметр розділення.

Розрахунок параметрів у відсотках здійснюється за формулами:

$$S_p = \sqrt{\frac{pq}{n}}, \text{ де } S_p - \text{середня квадратична похибка вибіркового середнього};$$

p – відсоток респондентів, що дали першу відповідь;

q – відсоток респондентів, що підтримали іншу відповідь;

n – обсяг вибірки.

$$\bar{P} = P \pm ZS_p \text{ де } \bar{P} - \text{середнє для генеральної сукупності};$$

P – параметр, що вивчається;

Z – параметр розділення.

Аналіз відмінностей – вид аналізу базується на гіпотезі, що дві сукупності не мають відмінностей між собою, а відмінності носять

випадковий характер. Результатом аналізу має стати рішення про можливість інтерполяції результату на всю сукупність.

Аналіз зв'язків – визначення ступеня впливу одного чинника на інший. Відповідає на питання, чи вплинуть зміни даного показника на інший.

Прийнято виділяти чотири типи зв'язків.

1. Немонотонний – означає, що присутність однієї змінної систематично пов'язана з іншою, але відсутня інформація про напрям цього зв'язку.

2. Монотонний – показує напрям зв'язку без визначення кількісних показників.

3. Лінійний – прямий зв'язок між чинниками, коли наявність даних про зміни одного автоматично дозволяє розрахувати інший.

$$y = a + bx,$$

де: y – змінна, яка оцінюється;

a – вільний член рівняння;

b – коефіцієнт регресії;

x – незалежна змінна, яка використовується для визначення залежної.

4. Криволінійний – описує зв'язок між змінними, що має складніший характер. Наприклад, всім відомі криві попиту і пропозиції. Зв'язок між змінними визначається такими характеристиками: наявність, напрям, сила зв'язку.

Метою **прогнозного аналізу** є прогнозування розвитку подій на основі наявної інформації. Основна складність даної процедури полягає в неможливості урахування всіх змінних, що можуть впливати на розвиток. Прогноз можна провести на основі евристичних методів, що засновані на суб'єктивних оцінках, інтуїції та досвіді особи, яка проводить дослідження, а також на основі статистичних методів розрахунку, які на практиці можуть бути й менш точними, ніж попередній спосіб.

Третій етап передбачає подання заключного звіту результатів дослідження викладеного у зрозумілому вигляді, що дозволить досягти остаточної мети дослідження й ухвалити стратегічне рішення.

Теми доповідей для заслуховування:

1. Аналіз сучасного стану на ринку статистичних пакетів.
2. Суть та організація статистичного зведення.
3. Візуалізація та графічне відображення даних.

Питання для контролю знань:

1. Надайте визначення термінам «аналіз даних», «спостереження», «експеримент», «імітація», «попередня обробка даних».
2. Зміст попередньої обробки даних.
3. Завдання попередньої обробки даних.
4. Методи обробки експериментальних даних.

5. Основні форми представлення результатів.
6. Варіанти представлення інформації.
7. Спосіб представлення результатів наукової роботи.
8. Методи статистичної обробки результатів експерименту.
9. Первинні методи математико-статистичного аналізу.
10. Вторинні методи математико-статистичного аналізу.
11. Значення понять «мода», «медіана», «вибіркове середнє», «розкид вибірки», «дисперсія».

Завдання для самостійної підготовки:

- I. Вивчити матеріал лекції.
- II. Підготувати доповідь за темою:
 1. Застосування t-критерію Стюдента (t-критерій).
 2. Застосування критерію Фішера (F-критерій).
 3. Застосування критерію Кохрена (G-критерій).
 4. Застосування критерію узгодженості Пірсона (χ^2 -критерій).
- III. Пройти тестування: тест 6.2. за посиланням
<https://forms.gle/2y69X4GYfxHF4XfY8>

ПРАКТИЧНЕ ЗАНЯТТЯ № 7

ОДНОФАКТОРНИЙ ДИСПЕРСІЙНИЙ АНАЛІЗ

Література:

1. Карагодова О.О., Рожок В.Д. Дослідження операцій: навч. посіб. К.: Центр учбової літератури, 2007. 256 с.
2. Бахрушин В.Є. Методи аналізу даних: навчальний посібник для студентів. Запоріжжя: КПУ, 2011. 268 с.
3. Перегуда О.В., Капустян О.А., Курилко О.Б. Статистична обробка даних: навч. посіб. Електронне видання, 2022. 103 с.
4. Рогальський Ф. Б., Курилович Я. Є., Цокурєнко А. А. Математичні методи аналізу економічних систем. Книга 1. Київ: Наукова думка, 2001, 435 с.
5. Рогальський Ф. Б., Цокурєнка А. А. Математичні методи аналізу економічних систем. Книга 2. Київ: Наукова думка, 2001, 423 с.
6. Крайчук О.В., Московська Г.К., Соколенко О.І. Теорія ймовірностей і математична статистика. Рівне, 2004.

Мета:

1. Навчальна – вивчити основи застосування факторного аналізу.
2. Виховна – виховувати професійні риси, почуття відповідальності; виховувати зацікавленість дисципліною, прагнення отримати нові знання самостійно.
3. Розвиваюча – прививати навички наукового підходу під час вирішення задач у сфері забезпечення пожежної безпеки.

Питання для обговорення:

1. Сутність моделі факторного аналізу, його основні завдання.
2. Класифікація методів факторного аналізу.
3. Визначення структури та статистичне дослідження моделі факторного аналізу.
4. Метод головних компонент.
5. Однофакторний аналіз.

Теоретична частина:

1. Сутність моделі факторного аналізу, його основні завдання.

Факторний аналіз — це багатовимірний метод, який застосовується для вивчення взаємозв'язків між значеннями змінних. Передбачається, що відомі змінні залежать від меншої кількості невідомих змінних і випадкової помилки.

Факторний аналіз – сукупність методів, які на основі реально існуючих зв'язків ознак (або об'єктів) дозволяють виявляти латентні (приховані) узагальнювальні характеристики організаційної структури та механізму розвитку досліджуваних явищ і процесів.

Факторний аналіз дозволяє вирішити дві важливі проблеми дослідника: описати об'єкт вимірювання всебічно та водночас компактно. За допомогою факторного аналізу можливе виявлення прихованих змінних факторів, що відповідають за наявність лінійних статистичних зв'язків між спостережуваними змінними.

Дві основні мети факторного аналізу:

- визначення взаємозв'язків між змінними, (класифікація змінних), тобто «об'єктивна R-класифікація»;
- скорочення числа змінних, необхідних для опису даних.

Завдання факторного аналізу наведено на рис.

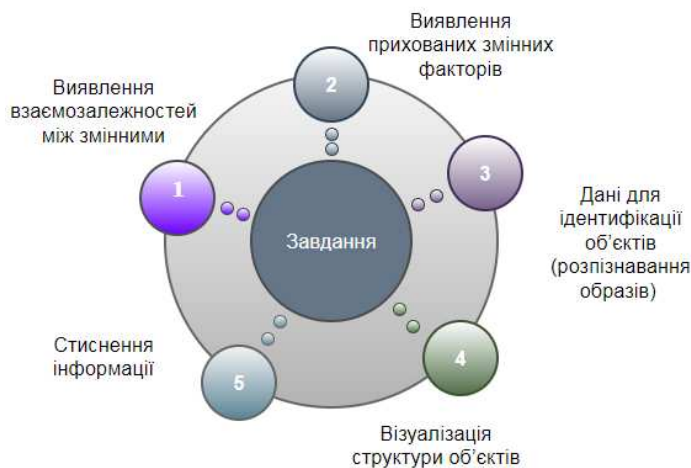


Рис. Завдання факторного аналізу

У ході аналізу в один фактор об'єднуються змінні, які сильно корелюють між собою. Унаслідок відбувається перерозподіл дисперсії між компонентами та створюється максимально проста та наочна структура факторів. Після об'єднання корельованість компонент між собою усередині кожного фактора буде вищою, ніж їх корельованість з іншими факторами. Ця процедура також дозволяє виділити латентні змінні, що особливо важливе для аналізу економічних систем. Аналізуючи оцінки, отримані за декількома шкалами, подібними між собою, і тими, які мають високий коефіцієнт кореляції, можна припустити, що існує деяка латентна змінна, за допомогою якої можна пояснити схожість отриманих оцінок. Таку латентну змінну називають фактором. Такий фактор впливає на численні показники інших змінних, що приводить до можливості та необхідності виділити його як найбільш загальний, більш високого порядку.

Окреслимо історію розвитку факторного аналізу:

- у 1901 р. виходить стаття англійського вченого К. Пірсона «На прямих і площинах, найбільш близьких до системи точок в просторі», у якій обговорювалась ідея головних осей;
- у 1904 р. опубліковано фундаментальну статтю Ч. Спірмена «Загальний інтелект, об'єктивно визначений та вимірний»;

- 40–50-ті роки – глибокі розробки американських статистиків і математиків: Л. Гуттмана, Г. Хотеллінга, Л. Терстоуна, К. Хользингера, С. Рао, англійських: С. Барта, Г. Томсона, Д. Лоулі, А. Максвелла та ін;
- у 60-ті роки відбувається розвиток методів факторного аналізу рішеннями Г. Кайзера, Р. Йорескога та ін;
- у 1960 р. виходить праця американського вченого Г. Хармана «Сучасний факторний аналіз».

2. Класифікація методів факторного аналізу.

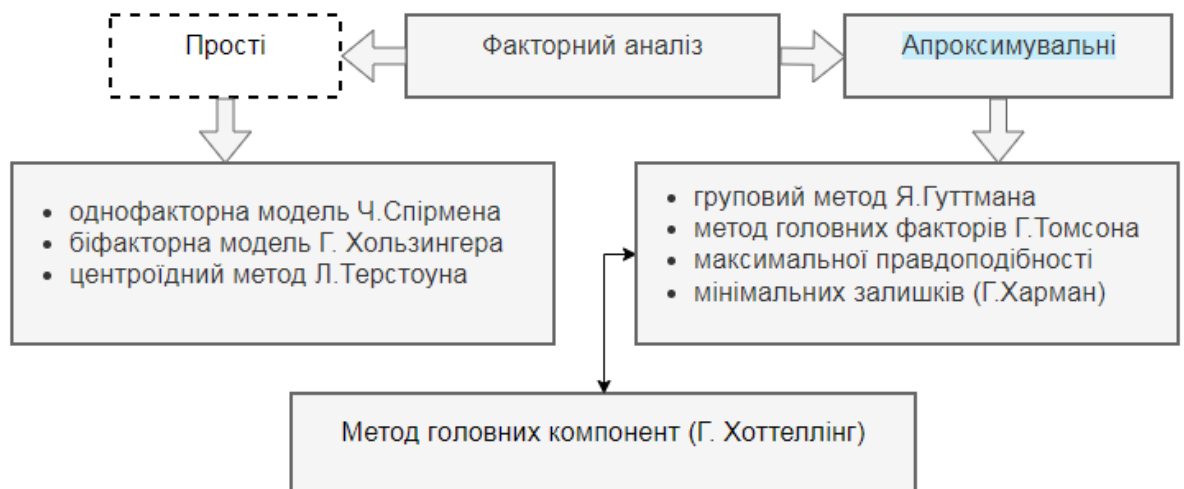


Рис. Класифікація методів факторного аналізу

Факторний аналіз розподіляють на такі види:

- *розвідувальний* — він здійснюється під час дослідження прихованої структури фактора без припущення про кількість факторів і їх навантажень;
- *конфірматорний* (підтверджувальний), призначений для перевірки гіпотез про кількість факторів і їх навантаженнях.

Практичне виконання факторного аналізу починається з перевірки його умов.

До обов'язкових умов факторного аналізу належать такі:

- усі ознаки мають бути кількісними;
- кількість спостережень має бути не менше ніж удвічі більше числа змінних;
- вибірка має бути однорідна;
- вхідні змінні мають бути розподілені симетрично;
- факторний аналіз здійснюється за корельованими змінними.

3. Визначення структури та статистичне дослідження моделі факторного аналізу.

Постановка завдання факторного аналізу. Нехай задана система змінних або ознак, значення яких відомі для кожного з N об'єктів. Уявімо вихідну інформацію у вигляді матриці розмірності $(n \times N)(n \times N)$. Передбачається, що кожен елемент матриці є результатом впливу деякого числа m гіпотетичних чинників і одного характерного. Для побудови моделі факторного аналізу слід провести стандартизацію вихідних даних.

$$y_{ij} = \frac{x_{ij} - \bar{x}_j}{s_j}, \quad s_j = \sqrt{\sum_{i=1}^N (x_{ij} - \bar{x}_j)^2}, \quad (8.1)$$

де y_{ij} — стандартизоване значення j -ї змінної на i -му об'єкті (безрозмірна величина);

s_j — середньоквадратичне відхилення j -ї ознаки (виправлене).

Основна модель факторного аналізу:

$$y_{ij} = a_{j1}f_{1i} + a_{j2}f_{2i} + \dots + a_{jm}f_{mi} + d_jv_{ji}, \quad j = \overline{1, n}; \quad r = \overline{1, m}; \quad i = \overline{1, N}, \quad (8.2)$$

де y_{ij} — нормоване значення j -го показника (змінної) j -го об'єкта дослідження;

f_{mi} — значення m -го загального фактора на i -му об'єкті дослідження;

v_{ji} — значення j -го характерного фактора на i -му об'єкті дослідження;

a_{jm} — ваговий коефіцієнт j -ї змінної на m -му загальному факторі або навантаження j -ї змінної на r -му загальному факторі;

d_j — навантаження або ваговий коефіцієнт j -ї змінної на j -му характерному факторі.

Матрична форма моделі факторного аналізу має такий вигляд:

$$\begin{aligned}
A &= \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1m} & 0 & 0 & \dots & 0 \\ a_{21} & a_{22} & \dots & a_{2m} & 0 & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & a_{nm} & 0 & 0 & \dots & 0 \end{pmatrix} + \\
+D &= \begin{pmatrix} 0 & 0 & \dots & 0 & d_1 & 0 & \dots & 0 \\ 0 & 0 & \dots & 0 & 0 & d_2 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 0 & 0 & 0 & \dots & d_n \end{pmatrix} = \\
= M &= \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1m} & d_1 & 0 & \dots & 0 \\ a_{21} & a_{22} & \dots & a_{2m} & 0 & d_2 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & a_{nm} & 0 & 0 & \dots & d_n \end{pmatrix}
\end{aligned}$$

Після цих уточнень модель можна записати в матричній формі: $Y = AF$. Або $Y = MF^*$, з врахуванням повної факторної матриці M :

$$\begin{aligned}
F &= \begin{pmatrix} f_{11} & \dots & f_{1N} \\ \dots & \dots & \dots \\ f_{m1} & \dots & f_{mN} \end{pmatrix} + V = \begin{pmatrix} v_{11} & \dots & v_{1N} \\ \dots & \dots & \dots \\ v_{m1} & \dots & v_{mN} \end{pmatrix} = \\
&= F^* = \begin{pmatrix} f_{11} & \dots & f_{1N} \\ \dots & \dots & \dots \\ f_{m1} & \dots & f_{mN} \\ v_{11} & \dots & v_{1N} \\ \dots & \dots & \dots \\ v_{m1} & \dots & v_{mN} \end{pmatrix}
\end{aligned}$$

Дисперсія ознаки y_j в факторному аналізі може бути подана як:

$$S_j^2 = h_j^2 + d_j^2 = 1$$

де S_j^2 — дисперсія ознаки y_j ;

h_j^2 — внесок у дисперсію ознаки всіх m загальних факторів;

d_j^2 — внесок у дисперсію ознаки y_j характерного фактора v_j .

Дисперсія характерного фактора може бути подана як:

$$d_j^2 = b_j^2 + c_j^2 = 1,$$

де c_j^2 — компонента дисперсії, пов'язана зі специфікою параметра;

b_j^2 — компоненти, пов'язаної з помилками вимірювань.

Компоненти дисперсії в факторному аналізі розраховують за формулами,

Компоненти дисперсії

Частка дисперсії	Позначення	Формула
Повна дисперсія	S_j^2	$h_j^2 + b_j^2 + c_j^2 = h_j^2 + d_j^2 = 1$
Надійність	r_j^2	$h_j^2 + b_j^2 = 1 - c_j^2$
Спільність	h_j^2	$1 - d_j^2$
Характерність	d_j^2	$b_j^2 + c_j^2 = 1 - h_j^2$
Специфічність	b_j^2	$d_j^2 - c_j^2$
Дисперсія помилки	c_j^2	$1 - r_j^2$

Фундаментальна теорема факторного аналізу. У виразі $Y = MF^*$ невідомі дві матриці з трьох: M і F^* . Вихідні дані матриці Y дозволяють отримати матрицю R . Це матриця коефіцієнтів парної кореляції, або кореляційна матриця. Також для відтворення матриці парних кореляцій можна використовувати матрицю M :

$$\begin{aligned} R &= MM^T = AA^T + DD^T; \\ R_h &= AA^T, DD^T = D^2. \end{aligned} \quad (8.5)$$

Таким чином, матриця коефіцієнтів парної кореляції, отримана з вихідних показників, може бути відтворена за допомогою матриці M :

$$R = R_h + D^2 \quad (8.6)$$

Подамо елемент матриці R у розгорнутому вигляді:

$$R = \frac{YY^T}{N}. \quad (8.7)$$

У матричній формі цей вираз можна записати так:

$$r_{jk} = \frac{1}{N}(y_{j1}y_{k1} + y_{j2}y_{k2} + \dots + y_{jN}y_{kN}). \quad (8.8)$$

Скориставшись формулою $Y = AF$, перетворимо R в редуковану матрицю:

$$R_h = \frac{1}{N}AF(AF)^T = \frac{1}{N}AFF^T A^T = A \frac{1}{N}FF^T A^T. \quad (8.9)$$

Позначимо $\frac{1}{N}FF^T = C$, відповідно: $R_h = ACA^T$. Цей вираз називають теоремою факторного аналізу. Якщо загальні фактори не корельовані між собою, то C буде одиничною матрицею, отже $R_h = AA^T$.

Методи обчислення спільності:

- метод квадрата коефіцієнта множинної кореляції;
- метод найбільшого коефіцієнта кореляції за рядком;
- метод оцінки середнього коефіцієнта кореляції за рядком;
- метод триад для оцінки h_j

$$h_j^2 = \frac{r_{jk}r_{jl}}{r_{kl}}, \quad (8.10)$$

де r_{jk} — коефіцієнт кореляції в рядку j , що має найбільше значення стохастичною зв'язку між ознакою Y_j і змінної Y_k ;

r_{kl} — коефіцієнт кореляції, який має найбільше значення, що характеризує зв'язок ознаки $Y_j \subset Y_l$

За методом першого центроїдного фактора обчислення h_j проводять таким чином:

$$\hat{h}_j^2 = \frac{(\sum_{k=1}^n r_{jk})^2}{\sum_{k=1}^n \sum_{l=1}^2 r_{kl}}. \quad (8.11)$$

Таким чином за обчисленими спільностями визначають головні фактори. Через значення повної дисперсії та спільності можна обчислити внесок остаточної частки дисперсії.

4. Метод головних компонент.

З числа методів, що дозволяють узагальнювати значення елементарних ознак, метод головних компонент виділяється простою логічною конструкцією; на його прикладі стають зрозумілими загальна ідея і цільові настанови численних методів факторного аналізу. Для виявлення найбільш значущих факторів і їх структури, найбільш виправданим є метод головних компонент. Сутність методу полягає в заміні корельованих компонентів некорельованими факторами. Іншою важливою характеристикою методу є можливість виокремлення найбільш інформативних головних компонент і виключення інших з аналізу, що спрощує інтерпретацію результатів. Переваги методу також у тому, що він є математично обґрунтованим методом факторного аналізу. За твердженням ряду дослідників, метод головних компонент не є методом факторного аналізу, оскільки не розщеплює дисперсію індикаторів на загальну й унікальну.

Метод головних компонент дає можливість за m вхідними ознаками виділити m головних компонент, або узагальнених ознак. Простір головних компонентів є ортогональним. Математична модель головних компонент базується на логічному допущенні, що значення безлічі взаємозалежних ознак породжують деякий загальний результат.

Метод головних компонент використовується для вивчення взаємозв'язків між досліджуваними показниками. За його допомогою можна виявляти приховані показники (фактори), які відповідають за наявність лінійних статистичних зв'язків (кореляцій) між ними. Крім того, визначення більш впливових за умов проведення досліджень факторів серед первинно обраних показників, а також виявлення статистичного зв'язку визначають обґрунтованість висновків щодо ефективності тих чи інших впливів на досліджувану систему.

Алгоритм застосування та реалізації методу головних компонент стосовно дослідження груп кредитних ризиків відображено на рис.

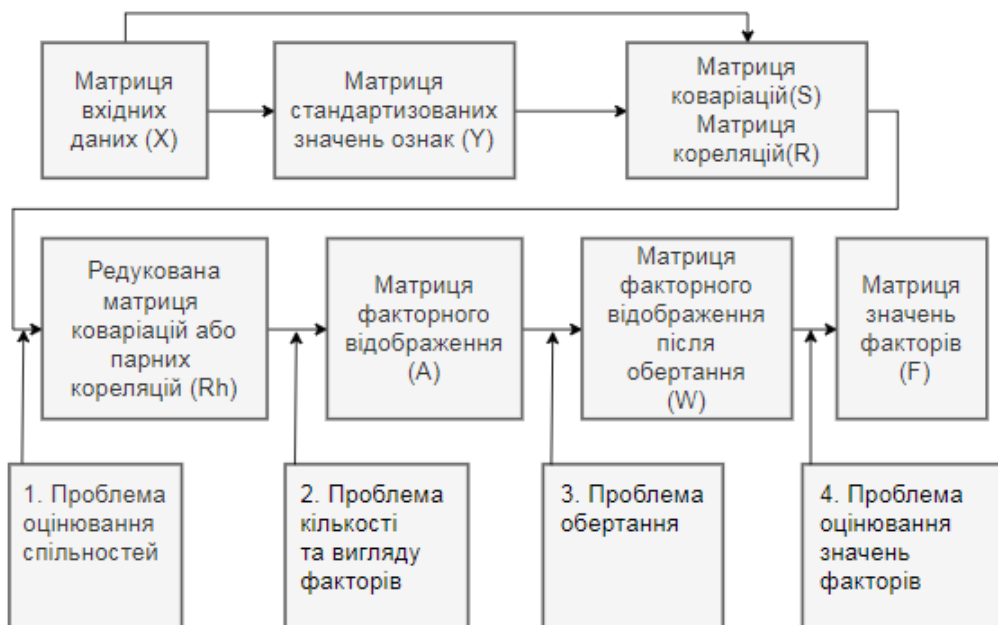


Рис. Загальна алгоритмічна схема реалізації методів факторного аналізу

5. Однофакторний аналіз.

Дисперсійний аналіз є сукупністю статистичних методів, призначених для перевірки гіпотез про зв'язок між певною ознакою та досліджуваними факторами, які не мають кількісного опису, а також для встановлення ступеня впливу факторів та їх взаємодії. Вперше цей метод було розроблено Р. Фішером в 1925 р.

Факторами називають контрольовані чинники, що впливають на кінцевий результат.

Рівнем фактора, або способом обробки, називають значення, що характеризують конкретний прояв цього фактора. Ці значення зазвичай подають у номінальній або порядковій шкалі вимірювань.

Значення вимірюваної ознаки називають **відгуком**.

Часто вихідні значення факторів вимірюють у кількісних або порядкових шкалах. Тоді постає проблема групування вихідних даних у ряди спостережень, що відповідають приблизно однаковим значенням фактора.

Якщо кількість груп взяти надмірно великою, то кількість спостережень у них може виявитися недостатньою для отримання надійних результатів.

Якщо її взяти надмірно малою, це може призвести до втрати суттєвих особливостей впливу досліджуваного фактора на систему. Вибір конкретного способу групування даних залежить від їх обсягу і характеру варіювання значень фактора.

Кількість і розміри інтервалів при однофакторному аналізі найчастіше визначають за принципом рівних інтервалів або за принципом

рівних частот. При багатофакторному аналізі застосовують три типи групування:

- групи з рівною кількістю спостережень;
- групи з різною кількістю спостережень;
- групи, кількості спостережень у яких відповідають певній пропорції.

При цьому існують певні особливості обробки даних, залежно від типу групування, які не розглядаються у цьому посібнику.

Основною метою однофакторного аналізу зазвичай є оцінка величини впливу конкретного фактора на досліджуваний відгук. Іншою метою може бути порівняння двох або декількох факторів один з одним з метою визначення різниці їх впливу на відгук, яку часто називають **контрастом факторів**. Попереднім етапом є перевірка нульової гіпотези про відсутність будь-якого впливу досліджуваного фактора (факторів), тобто гіпотези про те, що зміни значень ознаки в порівнюваних вибірках є випадковими, і всі дані належать до однієї генеральної сукупності.

Теми доповідей для заслуховування:

1. Поняття, типи і завдання факторного аналізу.
2. Класифікація факторів.
3. Систематизація факторів.
4. Поняття, типи і завдання факторного аналізу.
5. Метод головних факторів. Оцінювання факторів і задачі класифікації.
6. Метод головних компонент.
7. Приклади реалізації алгоритму методу головних компонент.

Питання для контролю знань:

1. У чому полягає сутність факторного аналізу?
2. Назвіть основні етапи факторного аналізу.
3. Які основні групи методів виділяють у факторному аналізі?
4. Які методи розрахунку спільностей ви знаєте?
5. Чим відрізняється редукована матриця від звичайної кореляційної матриці?
6. У чому відмінність методу головних компонент від методів факторного аналізу?
7. Як визначається необхідна та достатня кількість загальних факторів у методі головних факторів?
8. Які види обертання ви знаєте?
9. Яким чином здійснюється економічна інтерпретація отриманих чинників?
10. Який вигляд має лінійна модель методу головних компонент?
11. Назвіть основні поняття факторного аналізу.

12. Наведіть алгоритм методу головних факторів.
13. Як здійснюється розкладання дисперсії у факторному аналізі?
14. Назвіть критерії точності оцінки значень факторів.
15. Як здійснюється оцінювання рівня інформативності головних факторів?
16. Назвіть критерії вибору кількості факторів.

Завдання для самостійної підготовки:

- I. Вивчити матеріал лекції.
- II. Підготувати доповідь за темою:
 1. Теоретичні основи і принципова схема дисперсійного аналізу.
 2. Дисперсійний аналіз при групуванні даних за однією ознакою.
 3. Застосування дисперсійного аналізу для оцінки вірогідності різниці двох середніх.
 4. Дисперсійний аналіз при групуванні даних за двома ознаками.
- III. Пройти тестування: тест 7.2. за посиланням <https://forms.gle/jkDMNbC487uwU65W7>

ПРАКТИЧНЕ ЗАНЯТТЯ № 8

ДВОФАКТОРНИЙ ДИСПЕРСІЙНИЙ АНАЛІЗ

Література:

1. Карагодова О.О., Рожок В.Д. Дослідження операцій: навч. посіб. К.: Центр учбової літератури, 2007. 256 с.
2. Бахрушин В.Є. Методи аналізу даних: навчальний посібник для студентів. Запоріжжя: КПУ, 2011. 268 с.
3. Перегуда О.В., Капустян О.А., Курилко О.Б. Статистична обробка даних: навч. посіб. Електронне видання, 2022. 103 с.
4. Рогальський Ф. Б., Курилович Я. Є., Цокурєнко А. А. Математичні методи аналізу економічних систем. Книга 1. Київ: Наукова думка, 2001, 435 с.
5. Рогальський Ф. Б., Цокурєнка А. А. Математичні методи аналізу економічних систем. Книга 2. Київ: Наукова думка, 2001, 423 с.
6. Крайчук О.В., Московська Г.К., Соколенко О.І. Теорія ймовірностей і математична статистика. Рівне, 2004.

Мета лекції:

1. Навчальна – вивчити основи застосування дисперсійного аналізу.
2. Розвиваюча – ознайомитися з напрямками застосування дисперсійного аналізу.
3. Виховна – виховувати професійні риси, почуття відповідальності; виховувати зацікавленість дисципліною, прагнення отримати нові знання самостійно.

Питання для обговорення:

1. Задачі дисперсійного аналізу.
2. Двофакторний дисперсійний аналіз.

Теоретична частина:

1. Задачі дисперсійного аналізу.

В будь-якому експерименті середні значення досліджуваних величин змінюються у зв'язку зі зміною основних факторів (кількісних та якісних), що визначають умови досліду, а також і випадкових факторів. Дослідження впливу тих чи інших факторів на мінливість середніх є задачею дисперсійного аналізу.

Дисперсійний аналіз використовує властивість адитивності дисперсії випадкової величини, що обумовлено дією незалежних факторів. В залежності від числа джерел дисперсії розрізняють однофакторний та багатофакторний дисперсійний аналіз.

Дисперсійний аналіз особливо ефективний при вивченні кількох факторів. При класичному методі вивчення змінюють тільки один фактор, а решту залишають постійними. При цьому для кожного фактору

проводиться своя серія спостережень, що не використовується при вивченні інших факторів. Крім того, при такому методі досліджень не вдається визначити взаємодію факторів при одночасній їх зміні. При дисперсійному аналізі кожне спостереження служить для одночасної оцінки всіх факторів та їх взаємодії.

Дисперсійний аналіз полягає у виділенні й оцінюванні окремих факторів, що викликають зміну досліджуваної випадкової величини. При цьому проводиться розклад сумарної вибіркової дисперсії на складові, обумовлені незалежними факторами. Кожна з цих складових є оцінкою дисперсії генеральної сукупності. Щоб дати оцінку дієвості впливу даного фактору, необхідно оцінити значимість відповідної вибіркової дисперсії у порівнянні з дисперсією відтворення, обумовленою випадковими факторами. Перевірка значимості оцінок дисперсії проводять з допомогою критерію Фішера.

Коли розрахункове значення критерію Фішера виявиться меншим табличного, то вплив досліджуваного фактору немає підстав вважати значимим. Коли ж розрахункове значення критерію Фішера виявиться більшим табличного, то цей фактор впливає на зміни середніх. В подальшому ми вважаємо, що виконуються наступні припущення:

1. Випадкові помилки спостережень мають нормальний розподіл.
2. Фактори впливають тільки на зміну середніх значень, а дисперсія спостережень залишається постійною.

Фактори, що розглядаються в дисперсійному аналізі, бувають трьох родів:

- з випадковими рівнями, коли вибір рівнів проходить з безмежної сукупності можливих рівнів та супроводжується рандомізацією і рівні вибираються випадковим чином;
- з фіксованими рівнями;
- змішаного типу – частина факторів розглядається на фіксованих рівнях, але рівні решти вибираються випадковим чином.

Дисперсійний аналіз застосовується в різних формах в залежності від структури об'єкту, що досліджується; вибір відповідної форми є однією з головних труднощів в практичному застосуванні аналізу.

Основною метою дисперсійного аналізу, фундаментальна концепція якого була запропонована Фішером у 1920 р., є дослідження значущості відмінності між середніми декількох груп даних або змінних. Якщо порівнюються середні двох груп, дисперсійний аналіз дасть той же результат, що і звичайний t -критерій для незалежних або залежних вибірок. Проте використання дисперсійного аналізу має переваги особливо для малих вибірок.

У дисперсійному аналізі перевірка статистичної значущості відмінності між середніми декількох груп здійснюється на основі вибірових дисперсій. Ця перевірка проводиться за допомогою розбиття

загальної дисперсії (варіації) на частини, одна з яких обумовлена випадковою помилкою (тобто внутрішньогруповою мінливістю), а друга пов'язана з відмінністю середніх значень. Якщо ця відмінність значуща, нульова гіпотеза щодо існування відмінності між середніми значеннями відкидається на певному рівні значущості.

► Перевірка значимості оцінок дисперсій проводиться за критерієм Фішера.

► Необхідно, щоб виконувалися наступні припущення:

- 1) випадкові помилки спостережень мають нормальний розподіл;
- 2) фактори впливають тільки на зміну середніх значень, а дисперсія спостережень залишається постійною; експерименти рівнозначні.

Залежно від числа джерел дисперсії розрізняють однофакторний і багатфакторний дисперсійний аналіз.

2. Двофакторний дисперсійний аналіз.

Нехай необхідно визначити вплив двох факторів A і B на певну ознаку X . Для цього необхідно, щоб дослід здійснювався при фіксованих рівнях факторів A і B , а також їх одночасній дії на ознаку. При цьому дослід здійснюватимемо n разів для кожного з рівнів факторів A і B .

Позначимо через x_{ijk} конкретне значення ознаки X , якого вона набуває при i -му експерименті, j -му рівні фактора A і k -му рівні фактора B .

Результат експерименту зручно подати у вигляді таблиці, яка поділена на блоки, в кожному з яких враховується на певних рівнях факторів A і B їх вплив на конкретні значення ознаки $X=x_{ijk}$.

Виходячи з даних табл.,

$$\bar{x}_{ij} = \frac{\sum_{k=1}^n x_{ijk}}{n} \quad \text{є середнім значенням ознаки } X \text{ для кожного блока;}$$

$\bar{z}_j = \frac{\sum_{i=1}^n \sum_{k=1}^q x_{ijk}}{nq}, \quad j = \overline{1, p}$ є середнім значенням ознаки X за стовпцями;

$\bar{y}_i = \frac{\sum_{k=1}^n \sum_{j=1}^q x_{ijk}}{np}, \quad i = \overline{1, q}$ є середнім значенням ознаки X за рядками;

$$\bar{x} = \frac{\sum_{i=1}^n \sum_{j=1}^q \sum_{k=1}^p x_{ijk}}{npq} \quad \text{є загальною середньою ознакою } X;$$

$$S_1^2 = \frac{np \sum (\bar{y}_i - \bar{x})^2}{p-1} = \frac{Q_1}{p-1} \quad \text{є виправленою дисперсією, яка зумовлена впливом фактора } A \text{ на ознаку } X;$$

$$S_2^2 = \frac{nq \sum (\bar{z}_j - \bar{x})^2}{q-1} = \frac{Q_2}{q-1} \quad \text{є виправленою дисперсією, яка зумовлена впливом фактора } B \text{ на ознаку } X;$$

$$S_3^2 = \frac{\sum \sum (x_{ij} - \bar{z}_j - \bar{y}_i + \bar{x})^2}{(p-1)(q-1)} = \frac{Q_3}{(p-1)(q-1)} \quad \text{є виправленою дисперсією, яка зумовлена одночасним впливом на ознаку } X \text{ факторів } A \text{ і } B;$$

$$S_4^2 = \frac{\sum \sum \sum (x_{ijk} - \bar{x}_{jk})^2}{N - pq} = \frac{Q_4}{N - pq} \quad \text{є виправленою дисперсією, яка зумовлена впливом на ознаку } X \text{ інших, не головних факторів.}$$

Обчислюються спостережувані значення критерію

$$F_A^* = \frac{S_\sigma^2}{S_m^2}; \quad F_B^* = \frac{S_\sigma^2}{S_m^2}; \quad F_{AB}^* = \frac{S_\sigma^2}{S_m^2}.$$

При рівні значущості α визначають критичні точки:

$$F_{kp}(\alpha; k_4; k_1), \quad F_{kp}(\alpha; k_3; k_1), \quad F_{kp}(\alpha; k_2; k_2).$$

Якщо:

- 1) $F_A^* > F_{kp}(\alpha; k_4; k_1)$, то нульова гіпотеза про відсутність впливу фактора A відхиляється;
- 2) $F_B^* > F_{kp}(\alpha; k_3; k_1)$, то нульова гіпотеза про відсутність впливу фактора B відхиляється;
- 3) $F_{AB}^* > F_{kp}(\alpha; k_2; k_1)$, то нульова гіпотеза про відсутність спільного впливу факторів A B відхиляється.

Теми доповідей для заслуховування:

1. Теоретичні основи і принципова схема дисперсійного аналізу.
2. Дисперсійний аналіз при групуванні даних за однією ознакою.
3. Застосування дисперсійного аналізу для оцінки вірогідності різниці двох середніх.
4. Дисперсійний аналіз при групуванні даних за двома ознаками.

Питання для контролю знань:

1. Визначення понять: «аналіз», «дисперсія», «дисперсійний аналіз», «статистика», «проба», «фактор», «чинник», «гіпотеза», «адитивність».
2. Сутність однофакторного дисперсійного аналізу.
3. Сутність двофакторного дисперсійного аналізу.

Завдання для самостійної підготовки:

- I. Вивчити матеріал попередньої лекції.
- II. Пройти тестування: тест 8.2 за посиланням <https://forms.gle/uPSxMzk3EkHmxcY87>

ПРАКТИЧНЕ ЗАНЯТТЯ № 9

ПОБУДОВА ЛІНІЙНОЇ МОДЕЛІ. ЗАГАЛЬНІ ВІДОМОСТІ

Література:

1. Карагодова О.О., Рожок В.Д. Дослідження операцій: навч. посіб. К.: Центр учбової літератури, 2007. 256 с.
2. Бахрушин В.Є. Методи аналізу даних: навчальний посібник для студентів. Запоріжжя: КПУ, 2011. 268 с.
3. Перегуда О.В., Капустян О.А., Курилко О.Б. Статистична обробка даних: навч. посіб. Електронне видання, 2022. 103 с.
4. Рогальський Ф. Б., Курилович Я. Є., Цокурєнко А. А. Математичні методи аналізу економічних систем. Книга 1. Київ: Наукова думка, 2001, 435 с.
5. Рогальський Ф. Б., Цокурєнка А. А. Математичні методи аналізу економічних систем. Книга 2. Київ: Наукова думка, 2001, 423 с.
6. Крайчук О.В., Московська Г.К., Соколенко О.І. Теорія ймовірностей і математична статистика. Рівне, 2004.

Мета:

1. Навчальна – вивчити основи застосування кореляційного аналізу.
2. Розвиваюча – ознайомитися з напрямками застосування кореляційного аналізу.
3. Виховна – виховувати професійні риси, почуття відповідальності; виховувати зацікавленість дисципліною, прагнення отримати нові знання самостійно.

Питання для обговорення:

1. Визначення кореляційний аналіз.
2. Коефіцієнт кореляції.

Теоретична частина:

Дослідження носять комплексний характер. При контролі за ходом тренувального процесу оцінюють цілий ряд фізіологічних, біомеханічних, біохімічних та інших параметрів. При цьому часто виникає питання про взаємозв'язок ознак. Взаємозв'язок, при якому кожному значенню одного показника відповідає чітко визначене значення другого, називається функціональним.

Статистичний метод, який використовується для дослідження взаємозв'язків, називається кореляційним аналізом .

Кореляційний аналіз – це статистичне дослідження (стохастичної) залежності між випадковими величинами (англ. correlation – взаємозв'язок). У найпростішому випадку досліджують дві вибірки (набори даних), у загальному – багатовимірні комплекси (групи) геологічних параметрів або об'єктів.

Мета кореляційного аналізу – забезпечити отримання деякої інформації про одну змінну за допомогою іншої змінної. В випадках, коли можливе досягнення мети, говорять, що змінні корелюють. В загальному вигляді сприйняття гіпотези про наявність кореляції означає, що зміна значення змінної А відбудеться одночасно з пропорційною зміною значення В.

Кореляційний аналіз полягає у визначенні ступеня зв'язку між двома випадковими величинами X і Y. Дослідження характеру взаємозв'язку починається з побудови графічного зображення результатів вимірів в прямокутній системі координат, де кожна пара результатів буде відображатись точкою. Така графічна залежність називається діаграмою розсіяння або кореляційним полем.

Кореляційне поле відображає статистичний взаємозв'язок між результатами вимірів. Візуальний аналіз кореляційного поля дозволяє якісно оцінити форму, спрямованість і тісноту взаємозв'язку. Форма визначається по виду кореляційного поля: якщо через кореляційне поле можна провести пряму лінію – форма зв'язку лінійна, якщо ні – нелінійна.

Спрямованість взаємозв'язку визначається із залежності між результатами вимірів. Якщо у випадку покращення одного показника покращується другий – пряма залежність, спрямованість додатня. Збільшення одного результату викликає зменшення другого результату виміру – обернена залежність, спрямованість від'ємна.

Для оцінки статистичного взаємозв'язку, коли вимірювання проведено в шкалі відношень або інтервалів і форма взаємозв'язку лінійна, використовується коефіцієнт кореляції Браве-Пірсона, який обчислюється за наступною формулою:

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

де x_i, y_i - значення i -го результату виміру;

Для практичних розрахунків більш зручна наступна формула:

$$r = \frac{n \sum_{i=1}^n x_i y_i - \left(\sum_{i=1}^n x_i \right) \left(\sum_{i=1}^n y_i \right)}{\sqrt{\left[n \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i \right)^2 \right] \left[n \sum_{i=1}^n y_i^2 - \left(\sum_{i=1}^n y_i \right)^2 \right]}}$$

X, Y – середнє арифметичне показників вимірів X і Y ;

n – число вимірів (обсяг вибірки).

Знак коефіцієнта кореляції показує спрямованість зв'язку. Коефіцієнт кореляції зі знаком мінус показує обернену пропорційну залежність і навпаки – додатній знак коефіцієнта кореляції характеризує пряму залежність. Коли виміри виконані в шкалі порядку, то визначення взаємозв'язку показників проводять з використанням рангового коефіцієнта кореляції Спірмена, який обчислюють за наступною формулою:

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

де d_i – різниця рангів даної пари показників x_i і y_i ;

i – місце виміру в ранжованому ряді;

n – число вимірів.

Квадрат коефіцієнта кореляції називається коефіцієнтом детермінації:

$$d = r^2$$

Коефіцієнт детермінації визначає міру лінійної залежності, тобто ту частину загальної варіації одного показника, яка зумовлена варіацією іншого показника. Стверджують, що тільки d (%) взаємозв'язку спортивних результатів пояснюється їх взаємним впливом, а решта $(100 - d)$ % варіативності залежить від інших незумовлених факторів. При оцінці вірогідності коефіцієнтів взаємозв'язків потрібно встановити наступне: •

чи існує дійсно статистичний взаємозв'язок між двома явищами (параметрами, показниками)? • в яких вірогідних межах (границях) лежить істинне значення коефіцієнта кореляції в генеральній сукупності? Для лінійного парного коефіцієнта кореляції Браує-Пірсона це питання розв'язується за допомогою таблиці. Використання цієї таблиці дозволяє визначити значення коефіцієнта кореляції генеральної сукупності при заданому рівні значущості α і числі ступенів свободи ν . Розрахункове значення t-критерію Стюдента обчислюється за формулою:

$$t = \frac{r \sqrt{n-2}}{\sqrt{1-r^2}}$$

Порівнюючи значення розрахункового і табличного коефіцієнта кореляції можна з вірогідністю $P = (1 - \alpha) 100\%$ говорити про існування статистичного взаємозв'язку:

$r_{\text{роз}} < \Gamma_{\text{таб}}$ – взаємозв'язок не існує

$r_{\text{роз}} > \Gamma_{\text{таб}}$ – взаємозв'язок існує.

Оцінка вірогідності рангового коефіцієнта кореляції Спірмена виконується на основі t-критерію Стюдента, який розраховується за наступною формулою:

$$t = \frac{r_s \sqrt{n-2}}{\sqrt{1-r_s^2}}$$

де n – число пар вимірів

r_s - ранговий коефіцієнт кореляції Спірмена.

Коли розрахункове значення t-критерію Стюдента менше критичного $t_{\text{роз}} < t_{\text{таб}}$ то статистичний взаємозв'язок відсутній, і навпаки. Число ступенів свободи, як і у випадку лінійного парного коефіцієнта кореляції, також на два менше числа обсягу вибірки $\nu = n - 2$. Критичне значення t-критерію Стюдента визначається за таблицею

Мірою залежності між експериментальними наборами даних є числа – коефіцієнти зв'язку.

Головні завдання кореляційного аналізу:

- 1) оцінка за вибірковими даними коефіцієнтів кореляції;
- 2) перевірка значущості вибіркових коефіцієнтів кореляції або кореляційного відношення;
- 3) оцінка близькості виявленого зв'язку до лінійного;
- 4) побудова довірчого інтервалу для коефіцієнтів кореляції.

Визначення сили та напрямку взаємозв'язку між змінними є однією з важливих проблем аналізу даних. В загальному випадку для цього застосовують поняття кореляції.

Таким чином,

Кореляційний аналіз — це двовимірний статистичний метод вимірювання сили лінійного зв'язку між двома змінними та обчислення їх зв'язку. Простіше кажучи, кореляційний аналіз обчислює величину зміни однієї змінної через зміну іншої. Висока кореляція вказує на сильний зв'язок між змінними, тоді як низька кореляція означає, що змінні слабо залежать одна від одної. За допомогою кореляційного аналізу можна визначити зв'язки, закономірності, значущі зв'язки та тенденції між двома змінними або наборами даних.

Коефіцієнт кореляції — це числове значення, яке вказує на тип кореляції, тобто статистичний зв'язок між двома змінними. Значення коефіцієнта кореляції (r_s) коливається від + 1 до - 1 з точки зору сили зв'язку між змінними. Чим ближче значення коефіцієнта кореляції наближається до 0, тим слабкіше зв'язок між двома змінними. Напрямок зв'язку позначається знаком коефіцієнта кореляції; знак + вказує на пряму залежність, а знак - на зворотну залежність.

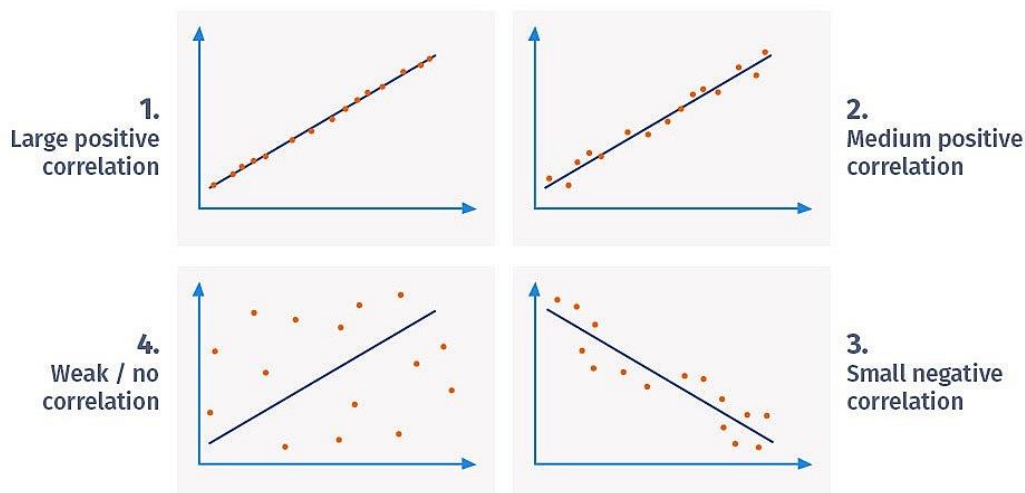
Коефіцієнт кореляції, а в загальному випадку кореляційна функція, дозволяють встановити степінь взаємозв'язку між змінними. Кореляція може бути лінійною або нелінійною в залежності від типу залежності, яка фактично існує між змінними. Досить часто на практиці розглядають тільки лінійну кореляцію (взаємозв'язок), але більш глибокий аналіз потребує використання для дослідження процесів нелінійних залежностей. Складну нелінійну залежність можна спростити, але знати про її існування необхідно для того, щоб побудувати адекватну модель процесу.

Кореляція між двома змінними може бути як позитивною кореляцією, так і негативною кореляцією або відсутність кореляції.

- **Позитивна кореляція:** Позитивна кореляція між двома змінними означає, що обидві змінні рухаються в одному напрямку. Збільшення однієї змінної призводить до збільшення іншої змінної і навпаки. Наприклад, якщо ви проводите більше часу на біговій доріжці, ви спалюєте більше калорій.

- **Негативна кореляція:** Від'ємна кореляція між двома змінними означає, що змінні рухаються в протилежних напрямках. Збільшення однієї змінної призводить до зменшення іншої змінної і навпаки. Наприклад, якщо ви збільшуєте швидкість транспортного засобу, час, необхідний для того, щоб дістатися до місця призначення, буде зменшуватися.

- **Слабка/нульова кореляція:** Немає кореляції, якщо одна змінна не впливає на іншу. Наприклад, не існує кореляції між кількістю років навчання в школі та кількістю літер у її імені.



Теми доповідей для заслуховування:

1. Кореляційний аналіз, мета і завдання.
2. Поняття кореляції.
3. Часткова та напівчасткова кореляції.
4. Властивості коефіцієнта кореляції.
5. Кореляційне поле.
6. Перевірка гіпотези про значущість коефіцієнта кореляції.
7. Автокореляція.
8. Кореляційна матриця.

Питання для контролю знань:

17. Визначення понять: «аналіз», «кореляція», «кореляційний аналіз», «коефіцієнт детермінації», «кореляційне поле».
18. Мета кореляційного аналізу.
19. Етапи кореляційного аналізу.
20. Коефіцієнт кореляції Браве-Пірсона.
21. t-критерій Стюдента.
22. Ранговий коефіцієнт кореляції Спірмена.
23. Позитивна кореляція.
24. Негативна кореляція.
25. Слабка/нульова кореляція.

Завдання для самостійної підготовки:

- I. Вивчити матеріал лекції.
- II. Підготувати доповідь за темою:
 1. Задачі регресійного аналізу.
 2. Регресійні моделі.

3. Базові (фундаментальні) положення класичного лінійного регресійного аналізу.

4. Класичні оцінки параметрів регресії методом найменших квадратів і їх властивості

5. Рекурентний алгоритм методу найменших квадратів.

6. Статистичний аналіз якості регресійної моделі.

III. Пройти тестування: тест 9.2. за посиланням <https://forms.gle/h3rijFgHtf3vhR5z7>

ПРАКТИЧНЕ ЗАНЯТТЯ № 10

ПОБУДОВА ЛІНІЙНОЇ МОДЕЛІ. ЗАГАЛЬНІ ВІДОМОСТІ

Література:

1. Літнарівч Р.М. Побудова і дослідження математичної моделі за джерелами експериментальних даних методами регресійного аналізу. Навчальний посібник, МEGУ, Рівне, 2011. 140 с.
2. Барковський В.В., Барковська Н.В., Лопатін О.К. Теорія ймовірностей та математична статистика : навч. посібник. 5-те вид. Київ : Центр учбової літератури, 2010. 424 с.
3. Блюмин С.Л., Суханов В.Ф., Чеботарьов С.В. Економічний факторний аналіз: монограф.. Липецьк: ЛЕГІ, 2004. 148 с.
4. Рогальський Ф.Б., Курилович Я.Є., Цокурєнка А.А. Математичні методи аналізу економічних систем. Книга 1. Київ: Наукова думка, 2001. 435 с.
5. Рогальський Ф.Б., Цокурєнка А.А. Математичні методи аналізу економічних систем. Книга 2. Київ: Наукова думка, 2001. 423 с.
6. Крайчук О.В., Московська Г.К., Соколенко О.І. Теорія ймовірностей і математична статистика. Рівне, 2004.

Мета:

1. Навчальна – вивчити основи застосування регресійного аналізу.
2. Розвиваюча – ознайомитися з напрямками застосування регресійного аналізу.
3. Виховна – виховувати професійні риси, почуття відповідальності; виховувати зацікавленість дисципліною, прагнення отримати нові знання самостійно.

Питання для обговорення:

1. Визначення поняття «регресійний аналіз».
2. Послідовність етапів регресійного аналізу.
3. Крива регресії.
4. Пряма регресії.
5. Криволінійні зв'язки.

Теоретична частина:

1. Визначення поняття «регресійний аналіз».

Регресійний аналіз — розділ математичної статистики, присвячений методам аналізу залежності однієї величини від іншої. На відміну від кореляційного аналізу не з'ясовує чи істотний зв'язок, а займається пошуком моделі цього зв'язку, вираженої у функції регресії.

Регресійний аналіз використовується в тому випадку, якщо відношення між змінними можуть бути виражені кількісно у виді деякої комбінації цих змінних. Отримана комбінація використовується для

передбачення значення, що може приймати цільова (залежна) змінна, яка обчислюється на заданому наборі значень вхідних (незалежних) змінних. У найпростішому випадку для цього використовуються стандартні статистичні методи, такі як лінійна регресія.

Основна особливість регресійного аналізу: при його допомозі можна отримати конкретні відомості про те, яку форму і характер має залежність між досліджуваними змінними.

2. Послідовність етапів регресійного аналізу.

Розглянемо коротко етапи регресійного аналізу.

1. Формулювання задачі. На цьому етапі формуються попередні гіпотези про залежність досліджуваних явищ.
2. Визначення залежних і незалежних (що пояснюють) змінних.
3. Збір статистичних даних. Дані повинні бути зібрані для кожної із змінних, включених в регресійну модель.
4. Формулювання гіпотези про форму зв'язку (проста або множинна, лінійна або нелінійна).
5. Визначення функції регресії (полягає в розрахунку чисельних значень параметрів рівняння регресії)
6. Оцінка точності регресійного аналізу.
7. Інтерпретація отриманих результатів. Отримані результати регресійного аналізу порівнюються з попередніми гіпотезами. Оцінюється коректність і правдоподібність отриманих результатів.
8. Передбачення невідомих значень залежної змінної.

За допомогою регресійного аналізу можливе вирішення задачі прогнозування і класифікації. Прогнозні значення обчислюються шляхом підстановки в рівняння регресії параметрів значень пояснюючих змінних. Вирішення задачі класифікації здійснюється таким чином: лінія регресії ділить всю множину об'єктів на два класи, і та частина множини, де значення функції більше нуля, належить до одного класу, а та, де вона менше нуля, - до іншого класу.

Розглянемо основні задачі регресійного аналізу: встановлення форми залежності, визначення функції регресії, оцінка невідомих значень залежної змінної.

Встановлення форми залежності.

Характер і форма залежності між змінними можуть утворювати наступні різновиди регресії:

- позитивна лінійна регресія (виражається в рівномірному зростанні функції);
- позитивна рівноприскорена зростаюча регресія;
- позитивна рівносповільнена зростаюча регресія;
- негативна лінійна регресія (виражається в рівномірному падінні

функції);

- негативна рівноприскорена спадаюча регресія;
- негативна рівносповільнена спадаюча регресія.

Проте описані різновиди зазвичай зустрічаються не в чистому вигляді, а в поєднанні один з одним. В такому разі говорять про комбіновані форми регресії.

Визначення функції регресії.

Друга задача зводиться до з'ясування дії на залежну змінну головних чинників або причин, за незмінних інших рівних умов, і за умови виключення дії на залежну змінну випадкових елементів. Функція регресії визначається у вигляді математичного рівняння того або іншого типу.

Оцінка невідомих значень залежної змінної.

Вирішення цієї задачі зводиться до вирішення задачі одного з типів:

- Оцінка значень залежної змінної усередині даного інтервалу вихідних даних, тобто пропущених значень; при цьому вирішується задача інтерполяції.
- Оцінка майбутніх значень залежної змінної, тобто знаходження значень поза заданим інтервалом вихідних даних; при цьому вирішується задача екстраполяції.

Обидві задачі вирішуються шляхом підстановки в рівняння регресії знайдених оцінок параметрів значень незалежних змінних. Результат вирішення рівняння є оцінкою значення цільової (залежною) змінної.

Розглянемо деякі припущення, на які спирається регресійний аналіз.

- Припущення лінійності, тобто передбачається, що зв'язок між даними змінними є лінійним. Так, в даному прикладі ми побудували діаграму розсіювання і змогли побачити явний лінійний зв'язок. Якщо ж на діаграмі розсіювання змінних ми бачимо явну відсутність лінійного зв'язку, тобто присутній нелінійний зв'язок, слід використовувати нелінійні методи аналізу.

- Припущення про нормальність *залишків*. Воно допускає, що розподіл різниці передбачених і спостережуваних значень є нормальним. Для візуального визначення характеру розподілу можна скористатися гістограмами.

- При використанні регресійного аналізу слід враховувати його основне обмеження. Воно полягає в тому, що регресійний аналіз дозволяє виявити лише залежності, а не зв'язки, що лежать в основі цих залежностей.

- Регресійний аналіз дає можливість оцінити міру зв'язку між змінними шляхом обчислення передбачуваного значення змінної на підставі декількох відомих значень.

Планові кількісні показники можливо розраховувати за допомогою декількох способів. Найбільш простим є «регресійний», тобто розрахунок шляхом екстраполяції від досягнутого рівня. Тією ж мірою

малообґрунтованою представляється орієнтація показників на результати найбільш успішно діючих філій чи опора на зобов'язання, взяті самими філіями.

Регресійний метод аналізу вирішує два основні завдання :

– визначають за допомогою рівнянь регресії аналітичного форму зв'язку між варіацією ознак X і Y ,

– встановлюють ступінь щільності зв'язку між ознаками.

Регресійні моделі бувають декількох видів:

- лінійна однофакторна регресійна модель (модель першого порядку);

- квадратична параболічна однофакторна регресійна модель (модель другого порядку) – характерна для багатьох процесів в текстильній промисловості;

- непараболічна однофакторна регресійна модель, перетворена в лінійну;

- параболічна однофакторна регресійна модель (поліноміальна модель любого порядку).

При обробці даних однофакторного експерименту здійснюються наступні операції:

1. Виключення даних, які різко відрізняються.

2. Перевірка гіпотези про нормальний розподіл випадкових величин.

3. Перевірка гіпотези про однорідність дисперсій в дослідних матриці.

4. Визначення середньої дисперсії вихідного параметру в дослідних матриці.

5. Визначення виду регресійної моделі.

6. Визначення коефіцієнтів регресії.

7. Визначення адекватності отриманого рівняння (за критерієм Фішера).

8. Визначення значимості коефіцієнтів регресії та їх довірчих інтервалів (за критерієм Стьюдента).

9. Визначення довірчих інтервалів середніх значень вихідного параметру при фіксованому значенні фактора.

10. Визначення довірчих інтервалів для індивідуальних значень вихідного параметру при кожному рівні фактора.

У регресійних моделях залежна (пояснюється) змінна Y може бути представлена у вигляді функції $f(X_1, X_2, \dots, X_k)$, де X_1, X_2, \dots, X_k — незалежні (що пояснюють) змінні, або фактори.

Зв'язок між змінною Y і k незалежними факторами X можна охарактеризувати функцією регресії $Y = f(X_1, X_2, \dots, X_k)$, яка показує, як буде в середньому значення змінної Y , якщо змінні X_i приймуть конкретні значення. Дана обставина дозволяє використовувати модель регресії не тільки для аналізу, але і для прогнозування економічних явищ.

Сформулюємо регресійну задачу для випадку однієї факторної ознаки.

Нехай є набір значень двох змінних: $Y = (y_1, y_2, \dots, y_n)$ — пояснюється мінлива і $X = (x_1, x_2, \dots, x_n)$ — пояснююча змінна, кожна з яких містить п спостережень. Нехай між змінними X і Y теоретично існує певна лінійна залежність

$$Y = f(X) = f(X_1, X_2, \dots, X_k) = \alpha + \beta x.$$

Це рівняння будемо називати «справжнім» рівнянням регресії.

Як вже відзначалось, зв'язок між взаємозалежними змінними величинами X і Y має математичний вираз у вигляді: $y = f(x)$. Наведене рівняння характеризує кореляційний характер зв'язків між X і Y і, водночас, відображає регресійну особливість, тобто є регресивним рівнянням. Це рівняння дозволяє визначити середні значення однієї ознаки (Y) за відповідними значеннями іншої ознаки (X). В правій частині даного рівняння коефіцієнт (f) є числовим значенням, яке фіксує, наскільки змінюється значення (Y) відносно значення (X). Це є коефіцієнт регресії або міра регресії. Він показує, на яке число збільшиться значення (Y) при збільшенні (X) на відповідну одиницю міри. При графічному зображенні від величини (f) залежить крутість розміщення лінії, що з'єднує одержані значення (Y).

Для всіх значень (x), що аналізуються, цей коефіцієнт в разі прямолінійної залежності завжди лишається постійним коефіцієнтом, що визначає міру збільшення або зменшення (y) щодо значень (x) або (x) щодо (y). Фактичне положення лінії (Y) відрізняється від лінії (X) на відповідну сталу величину (a), від якої залежить положення лінії $f(x)$ в системі координат. Отже, лінія регресії в системі координат в разі залежності $y = f(x)$ буде визначатись формулою прямої лінії $y = a + bx$, де (a) – сталий коефіцієнт, а (b) – коефіцієнт пропорційності, який зветься коефіцієнтом регресії або мірою регресії.

Коефіцієнт (a) означає рівень регресії, тобто її положення над віссю абсцис, який відраховується на початку координат від осі (x). Коефіцієнт (b) вказує наскільки збільшиться (або зменшиться) значення змінної ознаки (y), якщо значення (x) зміниться на одиницю виміру.

Мета регресійного аналізу:

- Визначення ступеня детермінованості варіації критеріальної (залежної) змінної предикторами (незалежними змінними).
- Пророкування значення залежної змінної за допомогою незалежної.
- Визначення внеску окремих незалежних змінних у варіацію залежної.

Регресійний аналіз не можна використовувати для визначення наявності зв'язку між змінними, оскільки наявність такого зв'язку і є передумова для застосування аналізу.

3. Крива регресії.

Регресія – форма зв'язку між випадковими величинами. Закон зміни математичного очікування однієї випадкової величини залежно від значень іншої. Розрізняють прямолінійну, криволінійну, ортогональну, параболічну, а також лінію і площину регресії.

Крива регресії Y на X є залежність умовного математичного очікування величини Y від заданого значення X :

$$m_{y/x} = \varphi(x, a, b, c, \dots),$$

де a, b, c, \dots — параметри рівняння регресії.

лінія регресії — емпірична в моделі аналітичного групування і теоретична в моделі регресійного аналізу. Емпірична лінія регресії представлена груповими середніми результативної ознаки, кожна з яких належить до відповідного інтервалу значень групувального фактора x_j . Теоретична лінія регресії описується певною функцією яку називають рівнянням регресії, а Y — теоретичним рівнем результативної ознаки.

На відміну від емпіричної, теоретична лінія регресії неперервна. Так, вважають, що маса дорослої людини в кілограмах має бути на 100 одиниць менша за її зріст у сантиметрах. Співвідношення між масою і зростом можна записати у вигляді рівняння: $y = 0,7x - 100$, де y — маса; x — зріст.

Безперечно, така форма зв'язку між масою та зростом людини надто спрощена. Насправді збільшення маси не жорстко пропорційне до збільшення зросту. Люди одного зросту мають різну масу, проте в середньому зі збільшенням зросту маса зростає.

Залежно від кількості змінних величин виділяють різні види регресійного аналізу. Якщо змінна величина завжди одна, то змінних може бути як одна, так і декілька. Виходячи з цього, виділяють два види регресійного аналізу: парний (простий) регресійний аналіз і регресійний аналіз на основі множинної регресії, або багатофакторний.

Парний регресійний аналіз – вид регресійного аналізу, що включає у себе розгляд однієї незалежної змінної величини, а багатофакторний – відповідно дві величини і більше.

Зважаючи на характер зв'язку, в регресійному аналізі можуть використовуватися лінійні та нелінійні функції. Для визначення характеру залежності та, відповідно, побудови рівняння регресії доцільно застосувати графічний метод, порівняння рівнобіжних рядів вихідних даних, табличний метод.

Так, графічний метод дає найбільш наочну картину розміщення крапок на графіку, завдяки чому можна виявити напрям і вид залежності між досліджуваними показниками: прямолінійна чи криволінійна.

Основне змістовне навантаження в рівнянні регресії несе коефіцієнт регресії. Найчастіше застосовуються лінійні рівняння або приведені до лінійного вигляду. Коефіцієнт регресії – це кутовий коефіцієнт у прямолінійному рівнянні кореляційного зв'язку. У лінійній функції рівняння регресії він показує на скільки одиниць в середньому зміниться

результативна ознака (y) при зміні факторної ознаки (x) на одиницю свого натурального виміру. Тобто, коефіцієнт регресії – це варіація y , яка припадає на одиницю варіації x . Коефіцієнт регресії має одиницю виміру результативної ознаки. За наявності прямого зв'язку коефіцієнт регресії є додатною величиною, а за зворотного зв'язку – від'ємною.

4. Пряма регресія.

В регресійному аналізі розрізняють рівняння парної (простої) та множинної (багатофакторної) регресії. Коли зв'язок із результативною ознакою y здійснюється з одним видом факторної ознаки x , то рівняння регресії має назву рівняння парної регресії. Якщо результативна ознака y пов'язана з декількома видами факторних ознак x , ($j=1-m$), то така залежність має назву рівняння множинної регресії. Обмежимося розглядом рівнянь парної регресії, як найбільш простим випадком зв'язку між ознаками, що достатньо широко використовується в статистичній практиці обстеження економічних явищ.

Відображення зв'язку між двома випадковими величинами x і y у вигляді залежності називають регресією y на x , і навпаки: y випадку кажуть про регресію x на y . Якщо лінії регресії є прямими, то регресію називають лінійною, в іншому випадку – нелінійною. Випадок прямої регресії є найпростішим, а тому найуживанішим в аналізі експериментальних даних.

Головною властивістю рівняння регресії є те, що вона (регресія) мінімізує суму квадратів (дисперсію) відхилень точок на лінії від експериментальних даних

5. Криволінійні зв'язки

Регресійний аналіз проводиться на основі побудованого рівняння регресії і визначає внесок кожної незалежної змінної y у варіацію досліджуваної (прогнозованої) залежної змінної величини.

Криволінійна регресія має місце тоді, коли із зміною (X) на відповідну одиницю виміру (Y) з кожним наступним збільшенням (X) змінюється не на сталу величину, а на якусь частинку більше або менше. Розмір цієї частинки регламентується ще одним коефіцієнтом «С», перемноженим на (x^2), тобто Cx^2 . Тоді формула лінійної регресії приймає вигляд $y = a + bx + Cx^2$, а крива на графіку буде відповідно згинатись від ординати (X) вгору (якщо значення Cx^2 буде додатнім) або вниз (значення Cx^2 – від'ємне). Криволінійність регресій може бути більш складною в порівнянні з наведеною. Тоді у математичну формулу регресії будуть включатись коефіцієнти, перемножені на наступні степені (x) – x^3 ; x^4 і т.д.

Відображення характеру залежності ознак (Y) і (X) відповідним математичним рівнянням зветься її апроксимацією (апроксимація – від лат. *approximatio* – зближення, тобто наближене зображення одних математичних об'єктів іншими, наприклад: ламаних ліній – кривими.)

Лінія регресії і відповідно до неї формула апроксимації обираються згідно, характером залежності одержаних експериментальних даних, який апробується за даними ранжированих рядів розподілу, а краще – за характером їх графічного виразу.

Звичайно за лінію вирівнювання приймають геометрично найбільш прості лінії: пряму, параболічну, логарифмічну, для яких за експериментальними даними складають рівняння зв'язку.

При цьому можуть бути такі випадки.

1. Якщо із збільшенням однієї ознаки спостерігається пропорційне збільшення або зменшення другої ознаки, за лінією вирівнювання обирається пряма лінія, тобто парабола першого порядку:

$$y = a + bx.$$

2. Якщо зміна залежного показника виражається плавною кривою з одним вигином, для вирівнювання беруть параболу другого порядку:

$$y = a + bx + cx^2.$$

3. В більш складних випадках – для кривих ~ образної форми, що мають два вигини, для вирівнювання експериментальних даних використовують параболу третього порядку:

$$y = a + bx + cx^2 + dx^3.$$

5. Якщо із збільшенням незалежної ознаки спостерігається уповільнене збільшення другої ознаки, застосовують логарифмічну криву:

6. Коли залежна ознака при збільшенні незалежної поступово зростає і це зростання переходить в пропорційне збільшення, для вирівнювання береться крива типу

В наведених формулах x – незалежна змінна; y – залежна змінна; a , b , c – постійні коефіцієнти, які підлягають визначенню.

Дуже часто ознаки біологічних об'єктів знаходяться в зворотній залежності одна від одної. Таку залежність обраховують рівняннями гіперболи:

Доцільно перед проведенням розрахунків рівнянь залежності графічно встановити відповідність характеру розподілу ознак тієї або іншої кривої розподілу.

Метод регресійного аналізу вважається найдосконалішим з усіх використовуваних нині нормативно-параметричних методів. Він широко застосовується для аналізу та встановлення рівня і співвідношень вартості продукції, яка характеризується наявністю одного або декількох техніко-економічних параметрів, що характеризують головні споживчі якості. Регресивний аналіз надає можливість знайти емпіричну форму залежності ціни від техніко-економічних параметрів товарів і виробів. При цьому він виступає в ролі цільової функції параметрів.

Метод регресійного аналізу особливо ефективний за умови здійснення розрахунків за допомогою сучасних інформаційних технологій і систем.

Вибір та обґрунтування функціонального виду регресії ґрунтується на теоретичному аналізі суті зв'язку. Нехай вивчається зв'язок між урожайністю та кількістю опадів. Надто мала і надто велика кількість опадів спричинюють зниження врожайності, максимальний її рівень можливий за умови оптимальної кількості опадів, тобто зі збільшенням факторної ознаки (опадів) урожайність спершу зростає, а потім зменшується. Залежність такого роду описується параболою $Y = a + bx + cx^2$.

Зауважимо, що теоретичний аналіз суті зв'язку, хоча й дуже важливий, лише окреслює особливості форми регресії і не може точно визначити її функціонального виду. До того ж у конкретних умовах простору і часу межі варіації взаємозв'язаних ознак x і y значно вужчі за теоретично можливі. І якщо кривина регресії невелика, то в межах фактичної варіації ознак зв'язок між ними досить точно описується лінійною функцією. Цим значною мірою пояснюється широке застосування лінійних рівнянь регресії.

Вивчення кореляційного зв'язку між ознаками починається з *регресійного аналізу*, який вирішує проблему встановлення форми зв'язку, або виду рівняння регресії, та визначення параметрів рівняння регресії.

В регресійному аналізі розрізняють рівняння *парної* (простої) та *множинної* (багатофакторної) регресії. Коли зв'язок із результативною ознакою у здійснюється з одним видом факторної ознаки x , то рівняння регресії має назву рівняння парної регресії. Якщо результативна ознака y пов'язана з декількома видами факторних ознак x_j ($j=1-t$), то така залежність має назву рівняння множинної регресії. Обмежимось розглядом рівнянь парної регресії, як найбільш простим випадком зв'язку між ознаками, що достатньо широко використовується в статистичній практиці обстеження економічних явищ.

Найбільш часто для характеристики кореляційного зв'язку між ознаками використовують такі види рівнянь парної регресії, або кореляційних рівнянь:

- а) лінійний $Y = a_0 + a_1x$;
- б) параболічний $Y = a_0 + a_1x^2$;
- в) гіперболічний $Y = a_0 + a_1 1/x$;
- г) степеневий $Y = a_0x^{a_1}$,

де a_0, a_1 - параметри рівнянь регресії, які підлягають визначенню.

Параметри a_j ($j=1-m$) в рівняннях регресії визначаються методом найменших квадратів, який запропоновано в XVIII ст. французьким математиком Лежандром. Цей метод найкращим чином відповідає кореляційній таблиці і припускає знаходження таких значень параметрів рівняння регресії, при яких сума квадратів відхилень табличних

(фактичних) значень результативної ознаки у від теоретичних Y за лінійною регресії була б мінімальною:

$$S = \sum(y - Y)^2 = \min$$

Функція S параметрів рівняння регресії a_j буде мінімальною тоді, коли виконуються необхідні умови знаходження екстремуму цієї функції - дорівнення нулю перших похідних функції за параметрами:

$$\partial S : \partial a_0 = 0; \partial S : \partial a_1 = 0$$

Із цих умов визначається система нормальних рівнянь для знаходження параметрів a_0 та a_1 .

У випадку лінійного виду рівняння регресій система нормальних рівнянь записується у вигляді:

$$\begin{aligned} \sum y &= n a_0 + a_1 \sum x \\ \sum xy &= a_0 \sum x + a_1 \sum x^2 \end{aligned}$$

де n - кількість одиниць сукупності (тобто заданих пар значень x і y).

Розв'язавши цю систему, знаходимо такі значення параметрів:

$$\begin{aligned} a_0 &= (\sum y \sum x^2 - \sum xy \sum x) : (n \sum x - \sum x \sum x) \\ a_1 &= (n \sum xy - \sum x \sum y) : (n \sum x^2 - \sum x \sum x) \end{aligned}$$

або

$$a_0 = y_{\text{сер}} - a_1 x_{\text{сер}}; a_1 = (x y_{\text{сер}} - x_{\text{сер}} y_{\text{сер}}) : (x^2_{\text{сер}} - x_{\text{сер}}^2)$$

де $x_{\text{сер}}$ - середня із добутку факторної ознаки на результативну.

Використавши рівняння регресії, можна знайти теоретичне значення Y для будь-якого значення факторної ознаки x .

У рівнянні параметр a_0 економічного змісту не має, а геометрично він відповідає значенню ординати ліній регресії Y при $x=0$. Параметр a_1 називається коефіцієнтом регресії і показує зміну результативної ознаки Y при зміні факторної ознаки x на одиницю; геометрично параметр a_1 відповідає куту нахилу (в радіанах) прямої лінії регресії до горизонтальної осі.

Для оцінки впливу факторної ознаки на результативну може розраховуватись коефіцієнт еластичності в середньому для усієї сукупності:

$$K_e = a_1 x_{\text{сер}} / y_{\text{сер}}$$

де $x_{\text{сер}}$, $y_{\text{сер}}$ - середні величини фактичних даних відповідно за факторною та результативною ознаками в цілому для сукупності.

Коефіцієнт еластичності показує, на скільки процентів у середньому зміниться результативна ознака при зміні факторної ознаки на 1%.

Прикладами використання лінійного рівняння регресії є такі залежності: між електроозброєністю праці (x) на 1 робітника і випуском готової продукції (y) для однорідних підприємств; між стажем роботи (x) та виробкою 1 робітника за зміну (y) тощо.

Для вибору виду рівняння регресії необхідно побудувати графік залежності фактичних даних $y=f(x)$ і за групуванням точок на графіку встановити візуально, до якого виду (лінійного чи нелінійного) можна віднести лінію регресії.

Теми доповідей для заслуховування:

1. Задачі регресійного аналізу.
2. Регресійні моделі.
3. Базові (фундаментальні) положення класичного лінійного регресійного аналізу.
4. Класичні оцінки параметрів регресії методом найменших квадратів і їх властивості
5. Рекурентний алгоритм методу найменших квадратів.
6. Статистичний аналіз якості регресійної моделі.

Питання для контролю знань:

1. Визначення понять: «аналіз», «регресія», «регресійний аналіз», «коефіцієнт регресії».
2. Основне завданням регресійного аналізу.
3. Що показує рівняння регресії. Основний вид.
4. Види регресійного аналізу.
5. Парний регресійний аналіз.
6. Припущення, на які спирається регресійний аналіз.
7. Лінійні та нелінійні функції.
8. Переваги графічного методу в регресійному аналізі.
9. Вид рівняння прямої.
10. Рівняння параболи.
11. Сутність аналізу на основі множинної регресії.
12. Багатофакторне рівняння множинної регресії.

Завдання для самостійної підготовки:

- I. Вивчити матеріал лекції.
- II. Підготувати доповідь за темою:
 1. Регресійний аналіз найпростіших поліноміальних моделей.
 2. Регресійний аналіз при неоднорідних і корельованих збуреннях.
 3. Регресійний аналіз в умовах мультиколінеарності.
 4. Вибір найкращої структури регресійної моделі.
 5. Регресійний аналіз в умовах похибок в регресорах.
- III. Пройти тестування: тест 10.2. за посиланням <https://forms.gle/hzz6VmAoRxWBU5Wr8>

ПРАКТИЧНЕ ЗАНЯТТЯ № 11 ПОБУДОВА НЕЛІНІЙНОЇ МОДЕЛІ. ЗАГАЛЬНІ ВІДОМОСТІ

Література:

1. Літнарівч Р.М. Побудова і дослідження математичної моделі за джерелами експериментальних даних методами регресійного аналізу. Навчальний посібник, МЕНУ, Рівне, 2011. 140 с.
2. Барковський В.В., Барковська Н.В., Лопатін О.К. Теорія ймовірностей та математична статистика : навч. посібник. 5-те вид. Київ : Центр учбової літератури, 2010. 424 с.
3. Блюмин С.Л., Суханов В.Ф., Чеботарьов С.В. Економічний факторний аналіз: монограф.. Липецьк: ЛЕГІ, 2004, 148 с.
4. Рогальський Ф.Б., Курилович Я.Є., Цокуренько А.А. Математичні методи аналізу економічних систем. Книга 1. Київ: Наукова думка, 2001, 435 с.
5. Рогальський Ф.Б., Цокуренько А.А. Математичні методи аналізу економічних систем. Книга 2. Київ: Наукова думка, 2001, 423 с.
6. Крайчук О.В., Московська Г.К., Соколенко О.І. Теорія ймовірностей і математична статистика. Рівне, 2004.

Мета:

1. Навчальна – вивчити загальні відомості про побудову нелінійної моделі.
2. Виховна – ознайомитися з методами побудови нелінійної моделі.
3. Розвиваюча – прививати навички наукового підходу у вирішенні питань побудови нелінійної моделі.

Питання для обговорення:

Загальні відомості про побудову нелінійної моделі.

Теоретична частина:

Завданням дослідження складних систем і процесів часто є перевірка наявності й встановлення типу зв'язку між незалежними змінними (**предикторами, факторами**), значення яких можуть змінюватися дослідником і мають певну задалегідь задану похибку, та залежною змінною (**відгуком**) г. Розв'язання таких завдань є предметом регресійного аналізу. Термін “Регресія” вперше був уведений Ф. Гальтоном наприкінці ХІХ ст. На практиці завдання регресійного аналізу зазвичай формулюють так: необхідно підібрати достатньо просту функцію, що в певному розумінні найкращим чином описує наявну сукупність емпіричних даних.

Загальна характеристика методів і задач регресійного аналізу.

Класичний регресійний аналіз включає методи побудови математичних моделей досліджуваних систем, методи визначення параметрів цих моделей і перевірки їх адекватності. Він припускає, що регресія є лінійною

комбінацією лінійно незалежних базисних функцій від факторів з невідомими коефіцієнтами (параметрами). Фактори й параметри є детермінованими, а відгуки - рівноточними (тобто мають однакові дисперсії) некорельованими випадковими величинами. Передбачається також, що всі змінні вимірюють у неперервних числових шкалах.

Звичайна процедура класичного регресійного аналізу є такою. Спочатку обирають гіпотетичну модель, тобто формулюють гіпотези про фактори, які суттєво впливають на досліджувану характеристику системи, і тип залежності відгуку від факторів. Потім за наявними емпіричними даними про залежність відгуку від факторів оцінюють параметри обраної моделі. Далі за статистичними критеріями перевіряють її адекватність.

При побудові регресійних моделей реальних систем і процесів вказані вище припущення виконуються не завжди. У більшості випадків їх невиконання призводить до некоректності застосування процедур класичного регресійного аналізу і потребує застосування більш складних методів аналізу емпіричних даних.

Постулат про рівноточність і некорельованість відгуків не є обов'язковим. У випадку його невиконання процедура побудови регресійної моделі певною мірою змінюється, але суттєво не ускладнюється.

Більш складною проблемою є вибір моделі та її незалежних змінних. У класичному регресійному аналізі припускають, що набір факторів задається однозначно, всі суттєві змінні наявні в моделі й немає ніяких альтернативних способів обрання факторів. На практиці це припущення не виконується. Тому виникає необхідність розробки формальних та неформальних процедур перетворення й порівняння моделей. Для пошуку оптимальних формальних перетворень використовують методи факторного та дискримінантного аналізу. На сьогодні розроблено комп'ютеризовані технології послідовної побудови регресійних моделей.

Фактори в класичному регресійному аналізі вважають детермінованими, тобто вважається, що дослідник має про них всю необхідну інформацію з абсолютною точністю. На практиці це припущення часто не виконується. Відмова від детермінованості незалежних змінних зумовлює необхідність застосування моделей кореляційного аналізу. В окремих випадках можна використовувати компромісні методи **конфлюентного аналізу**, які передбачають можливість нормально розподіленого та усіченого розкиду значень факторів. Якщо ця умова виконується, побудову моделі можна звести до багаторазового розв'язування регресійної задачі.

Відмова від припущення про детермінованість параметрів моделей у регресійному аналізі призводить до суттєвих ускладнень, оскільки порушує його статистичні основи. Але на практиці це припущення виконується не завжди. У деяких випадках можна вважати параметри випадковими величинами із заданими законами розподілу. Тоді як оцінки

параметрів можна брати їх умовні математичні сподівання для відгуків, що спостерігалися. Умовні розподіли та математичні сподівання розраховують за узагальненою формулою Байєса, тому відповідні методи називають **байєсівським регресійним аналізом**.

Регресійні моделі часто використовують для опису процесів, що розвиваються у часі. У певних випадках це зумовлює необхідність переходу від випадкових значень відгуків до випадкових послідовностей, випадкових процесів або випадкових полів. Однією з поширених і найпростіших моделей такого типу є **модель авторегресії**, згідно з якою відгук залежить не тільки від факторів, але також і від часу. Якщо останню залежність можна виявити, то проблема зводиться до стандартної задачі побудови регресії для модифікованого відгуку. В інших випадках необхідно використовувати більш складні прийоми.

Процедури класичного регресійного аналізу припускають, що закон розподілу відгуків є нормальним. Проте на практиці найчастішими є випадки, коли цей закон невідомий чи відомо, що він не є нормальним. Їх дослідження зумовило виникнення непараметричного регресійного аналізу, який не передбачає необхідності попереднього задання функції розподілу.

Важливою проблемою, яка виникає при оцінюванні параметрів регресійних моделей, є наявність грубих помилок серед набору аналізованих даних. Ці помилки можуть виникати внаслідок неправильних дій дослідника, збоїв у роботі апаратури, неконтрольованих короткотривалих сильних зовнішніх впливів на досліджувану систему тощо. У таких випадках

використовують два підходи, що дають змогу зменшити вплив грубих помилок на результати аналізу. У першому з них розробляють критерії та алгоритми пошуку помилкових даних. Потім ці дані відкидають. У другому підході розробляють алгоритми аналізу, які є нечутливими до наявних помилкових даних (алгоритми робастного оцінювання параметрів).

Одним з основних постулатів класичного регресійного аналізу є припущення, що найкращі оцінки параметрів можна одержати, використовуючи метод найменших квадратів. На практиці оцінки, одержані за допомогою цього методу, часто бувають недостатньо точними і містять великі похибки. Причиною цього може бути структура регресійної моделі. Якщо вона є лінійною комбінацією експонент або поліномом високого степеня, то це призводить до поганої зумовленості матриці системи нормальних рівнянь і нестійкості оцінок параметрів. Підвищення стійкості оцінок можна досягти шляхом відмови від вимоги щодо їх незміщеності. Розвиток цього напрямку досліджень призвів до виникнення гребеневого, або рідж-регресійного аналізу.

Найчастіше задачу побудови регресійної моделі формулюють так. Необхідно знайти функцію заданого класу, для якої функціонал:

$$F(\alpha) = \sum_{i=1}^n (z_i(\alpha, X) - y_i)^2 \rightarrow \min. \quad (1)$$

У цьому виразі $z_i(\alpha, X)$ - значення функції, що апроксимує залежність, в i -ї точці, y_i - відповідне значення емпіричної залежності, α - вектор параметрів, які треба знайти, X - вектор незалежних змінних. Одержану функцію $z(\alpha, X)$ називають **(середньоквадратичною) регресійною моделлю**. Метод її пошуку, який базується на застосуванні критерію (1), називають методом найменших квадратів.

Іноді замість функціонала (1) для визначення параметрів регресійних моделей розв'язують задачі мінімізації інших функціоналів, зокрема:

$$F(\alpha) = \sum_{i=1}^n |z_i(\alpha, X) - y_i| \rightarrow \min;$$

$$F(\alpha) = \max |z_i(\alpha, X) - y_i| \rightarrow \min.$$

Одержувані при цьому регресійні моделі називають, відповідно, **середньоабсолютними (медіанними)** та **мінімаксними**. Ці моделі найчастіше використовують при побудові робастних алгоритмів регресійного аналізу, але їх практичне застосування обмежується поганою збіжністю таких алгоритмів.

Апроксимуючу функцію у випадку однієї незалежної змінної (моделі

$$z(x) = \sum_{j=0}^M \alpha_j x^j,$$

простої регресії) часто шукають у вигляді полінома оберненого

$$z(x) = \frac{1}{\sum_{j=0}^M \alpha_j x^j},$$

полінома експоненціальних або показникових

$z = ae^x$ чи $z = ab^x$, степеневій функції $z = ax^b$, лінійно-логічній функції

$z = \alpha_1 + \alpha_2 x + \alpha_3 \ln x$, тригонометричного ряду Фур'є тощо. За наявності

декількох незалежних змінних (моделі множинної регресії) найчастіше використовують функції, лінійні як за параметрами, так і за незалежними

$$z = \alpha_0 + \sum_{i=1}^p \alpha_i x_i,$$

змінними а також поліноміальні моделі, що є лінійними за параметрами, але нелінійними за незалежними змінними:

$$z = \alpha_0 + \sum_{i=1}^p \alpha_i x_i + \sum_{\substack{i,j=1 \\ i \geq j}}^p \alpha_{ij} x_i x_j + \sum_{\substack{i,j,k=1 \\ i \geq j \\ i \geq k}}^p \alpha_{ijk} x_i x_j x_k + \dots$$

Останні відповідають розкладу функції відгуку в ряд Тейлора. Проте можливе й використання для апроксимації інших видів залежностей.

Регресійні моделі називають **лінійними** або **нелінійними**, якщо вони є, відповідно, лінійними або нелінійними за параметрами. При цьому визначення "лінійна" часто опускають. Значення найвищого степеня предиктора в поліноміальних моделях називають **порядком моделі**.

Наприклад:

$$z = \alpha_0 + \alpha_1 x + \alpha_2 x^2 + \alpha_3 x^3 + \varepsilon, \quad (2)$$

де ε - похибка моделі, ε лінійною моделлю третього порядку.

Вибір типу регресійної моделі є нетривіальним завданням. Для моделей, що містять одну незалежну змінну, рекомендують спочатку нанести наявні емпіричні дані на графік. Це дає можливість визначити наявність чи відсутності залежності між досліджуваними величинами, а також зробити певні припущення про тип залежності.

На рис. 1 як приклади наведено певні набори емпіричних точок, для яких потрібно побудувати регресійні моделі. З наведених графіків видно, що ці моделі доцільно будувати у вигляді лінійної, квадратичної та експоненціальної функцій, відповідно. Але, як правило, визначення типу моделі за графіком емпіричних даних є не настільки очевидним, тому зазвичай доводиться перевіряти декілька варіантів моделі і вибирати кращий з них за певними критеріями.

Часто як попередній етап регресійного аналізу рекомендують за допомогою методів кореляційного аналізу перевіряти наявність значущого зв'язку між досліджуваними змінними. Але при цьому слід урахувати, що звичайні методи кореляційного аналізу дають змогу перевіряти лише гіпотезу про наявність лінійного зв'язку. Якщо зв'язок є, але він нелінійний, висновки, отримані за допомогою кореляційного аналізу, можуть бути помилковими.

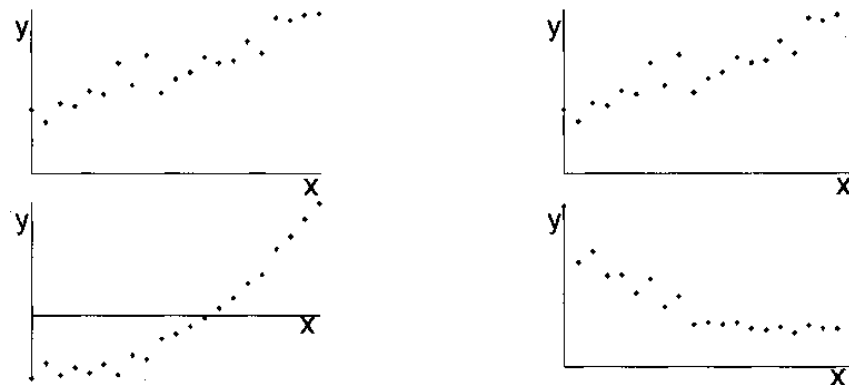


Рис. 1. Приклади наборів даних, для яких треба побудувати регресійні моделі

Важливою особливістю регресійних моделей є те, що їх не можна застосовувати поза межами тієї області значень вихідних параметрів, для якої вони були побудовані. При використанні регресійних моделей типу полінома, оберненого полінома, тригонометричного ряду та деяких інших слід враховувати, що, збільшуючи кількість членів ряду, можна одержати скільки завгодно близькі до нуля значення функціоналів (1). Проте це не завжди свідчить про якість апроксимації, оскільки ці функціонали не дають інформації про ступінь наближення моделі до емпіричної залежності у проміжках між наявними точками.

Іншою проблемою може бути наявність декількох локальних екстремумів функціоналів (1). У таких випадках необхідно враховувати, що більшість стандартних алгоритмів дає можливість знаходити лише локальні, а не глобальні екстремуми функціоналів, і результат мінімізації залежать від вибору початкових умов пошуку. Це часто зумовлює необхідність встановлення додаткових критеріїв вибору моделі, серед яких можуть бути як формальні критерії їх адекватності, так і неформальні критерії, що ґрунтуються на сукупності відомих даних про об'єкт дослідження.

Поліноміальні регресійні моделі, як правило, є формальними. Їх використовують для опису систем і процесів, теорію яких розроблено недостатньо. При цьому спираються на відомі властивості ряду Тейлора для аналітичних функцій. Більш цікавими для дослідників зазвичай є математичні моделі, які відображають структуру та зв'язки у системах, сутність і механізми процесів, що відбуваються у них. Якщо теоретичні основи досліджуваних систем і процесів достатньо розроблені, часто постає проблема визначення окремих параметрів моделі за наявними емпіричними даними. Для її вирішення у багатьох випадках можна використовувати формальні процедури регресійного аналізу.

На практиці часто доводиться користуватися нелінійними за параметрами та багатовимірними моделями. Під багатовимірними тут розуміють моделі, що розглядають декілька відгуків.

Математична модель для повного факторного експерименту

У разі повного факторного експерименту функція відгуку для параметра Y залежно від k кодованих факторів шукається у вигляді відрізка - ряду Тейлора

$$y = a_0 + a_1 x_1 + a_2 x_2 + \dots + a_k x_k \quad (3)$$

Наша мета - знайти за результатами дослідів, поставлених у всіх точках повного факторного експерименту, значення невідомих коефіцієнтів в математичній моделі (3).

Коефіцієнти регресії цієї моделі обчислюються за формулами:

$$a_i = \left(\sum_{j=1}^n x_{ij} y_{jcp} \right) / N, \quad i = 0, 1, 2, \dots, k$$

де номер стовпця $i = 0$ належить до нульового стовпця матриці планування, який складається з одних елементів +1. У випадку $k = 2$ маємо:

$$\begin{aligned} a_0 &= (y_1 + y_2 + y_3 + y_4) / 4 \\ a_1 &= (-y_1 - y_2 + y_3 + y_4) / 4 \\ a_2 &= (-y_1 + y_2 - y_3 + y_4) / 4 \end{aligned}$$

Ви бачите, що завдяки кодуванню факторів розрахунок коефіцієнтів a_i перетворився у просту арифметичну процедуру. Тепер у нас є все

необхідне, щоб знайти невідомі коефіцієнти лінійної моделі залежно від вхідних факторів у натуральному вимірі

$$y = b_0 + b_1 z_1 + b_2 z_2$$

Для отримання останнього рівняння достатньо скористатися формулами переходу від кодованих змінних до натуральних. Коефіцієнти при незалежних змінних вказують на ступінь впливу факторів. Чим більше чисельна величина коефіцієнта, тим більший вплив він надає. Якщо коефіцієнт має знак плюс, то зі збільшенням значення фактора параметр оптимізації збільшується, а якщо мінус, то зменшується. Величина коефіцієнта відповідає внеску даного фактора у величину параметра оптимізації при переході фактора з нульового рівня на верхній або нижній.

Внесок i -го фактора в загальний вплив на вихідну величину зручно оцінити значенням a_i , яке визначає змінну Y при переході цього фактору від нижнього до верхнього рівня. Внесок, визначений таким чином, отримав назву ефекту фактора (іноді його називають основним або головним ефектом). Він чисельно дорівнює подвоєному коефіцієнту. Для якісних факторів, варійованих на двох рівнях, основний рівень не має фізичного сенсу. Тому поняття «ефект фактора» є тут природним. Плануючи експеримент, на першому кроці ми прагнемо отримати лінійну модель. Однак у нас немає гарантії, що в обраних інтервалах варіювання процес описується лінійною моделлю. Існують способи перевірки придатності лінійної моделі.

Один з найпоширеніших видів нелінійності пов'язаний з тим, що ефект одного чинника залежить від рівня, на якому знаходиться інший фактор. У цьому випадку говорять, що має місце **ефект взаємодії** двох факторів. Повний факторний експеримент дозволяє кількісно оцінювати ефекти взаємодії. Для цього треба, користуючись правилом множення стовпців, отримати стовпець твори двох факторів. При обчисленні коефіцієнта, відповідного ефекту взаємодії, з новим вектор-стовпцем можна поводитися так само, як з вектор-стовпцем будь-якого фактора.

Для повного факторного експерименту матриця планування з урахуванням ефекту взаємодії продуктів представлена в табл. 4. Дуже важливо, що при додаванні стовпців ефектів взаємодій всі розглянуті властивості матриць планування зберігаються.

Таблиця 4 - Матриця планування експерименту 2^2 з ефектом взаємодій

Номер дослідів					x
1	1	1	1	1	+
2	1	1	1	1	-
3	1	1	1	1	-
4	1	1	1	1	+

Математична модель з урахуванням взаємодії виглядає наступним чином:

$$y = a_0 + a_1 x_1 + a_2 x_2 + a_{12} x_1 x_2 \quad (7)$$

Коефіцієнт взаємодії a_{12} обчислюється звичайним шляхом:
 $a_{12} = (y_1 - y_2 - y_3 + y_4)/4$ (8)

Стовпці x_1 і x_2 задають вхідні параметри моделі. За ним безпосередньо визначаються умови дослідів, стовпці x_0 і $x_1 x_2$ служать тільки для розрахунку.

Звертаємо вашу увагу на те, що при оптимізації ми прагнемо зробити ефекти взаємодії можливо меншими. У задачах інтерполяції, навпаки, їх виявлення часто важливо і цікаво.

Приклад побудови математичної моделі з двома факторами

Досліджувався вихід продукту реакції залежно від температури і тиску. Матриця планування в природному вигляді наведена в табл. 5. Кожен дослід дублювався двічі

Таблиця 5. Матриця планування і результати експерименту впливу температури z_1 і тиску z_2 на вихід продукту реакції Y .

i	$z_1 =$	$z_2 =$	$y_1, \%$	$y_2, \%$	$y_{\text{ср}}, \%$
1	140	28	83	87	85
2	140	32	75	75	75
3	160	28	77	83	80
4	160	32	71	73	72

Перейдемо до кодованим змінним. У даному випадку:

$$z_{1\text{max}} = 160, z_{1\text{min}} = 140, \quad z_{2\text{max}} = 32, z_{2\text{min}} = 28$$

$$Dz_1 = (z_{1\text{max}} - z_{1\text{min}})/2 = (160 - 140)/2 = 10$$

$$Dz_2 = (z_{2\text{max}} - z_{2\text{min}})/2 = (32 - 28)/2 = 2$$

Матриця планування в кодованому вигляді представлена в табл. 6.

Таблиця 6. Матриця планування експерименту вплива температури x_1 і тиску x_2 на вихід продукту реакції Y в кодованому вигляді.

i	x_1	x_2	$x_1 x_2$	x_1^2	x_2^2	y_1	y_2	y_c
1	1	-	-	+	8	8	8	
	1	1	1	3	7	5	8	
2	1	-	+	-	7	7	7	
	1	1	1	5	5	5	7	
3	1	+	-	-	7	8	8	
	1	1	1	7	3	0	8	
4	1	+	+	+	7	7	7	
	1	1	1	1	3	2	7	

Розрахунок дає:

$$a_0 = (85 + 80 + 75 + 72)/4 = 78$$

$$a_1 = (-85 + 80 - 75 + 72)/4 = -2,0$$

$$a_2 = (-85 - 80 + 75 + 72)/4 = -4,5$$

Шукана лінійна модель у кодованому вигляді має вигляд:

$$y = 78 - 2,0x_1 - 4,5x_2$$

Математична модель з урахуванням взаємодії має вигляд:

$$y = 78 - 2,0x_1 - 4,5x_2 + 0,5x_1x_2$$

де останній член, коефіцієнт взаємодії a_{12} , обчислювався в відповідності з формулою (4) і даними табл. 6:

$$a_{12} = (85 - 80 - 75 + 72)/4 = 0,5$$

Він визначає вплив першого фактора при наявності другого фактору і, навпаки, вплив другого чинника при наявності першого.

Теми доповідей для заслуховування:

1. Регресійний аналіз найпростіших поліноміальних моделей.
2. Регресійний аналіз при неоднорідних і корельованих збуреннях.
3. Регресійний аналіз в умовах мультиколінеарності.
4. Вибір найкращої структури регресійної моделі.
5. Регресійний аналіз в умовах похибок в регресорах.

Питання для контролю знань:

1. Яким є основне завдання регресійного аналізу?
2. У чому полягають основні припущення класичного регресійного аналізу?
3. Якою є звичайна процедура класичного регресійного аналізу?
4. Як формулюється задача побудови регресійної моделі?
5. Що таке «коефіцієнти регресії»?
6. Отримання коефіцієнтів регресії при побудові лінійної регресійної моделі.
7. Які моделі називають лінійними? Що називають порядком регресійної моделі?
8. Чому регресійні моделі не рекомендують використовувати поза межами тієї області значень вихідних параметрів, для якої вони побудовані?
9. Для заданого набору даних побудувати однофакторну лінійну регресійну модель і перевірити її адекватність.
10. Які функціонали використовують для визначення параметрів регресійних моделей? У чому полягають переваги й недоліки різних типів таких функціоналів?
11. Якими є основні типи функцій, що використовуються для побудови однофакторних регресійних моделей?

Завдання для самостійної підготовки:

- I. Пройти тестування: тест 11.2. за посиланням <https://forms.gle/VtPMSjGGKWzAR3pXA>
- II. Підготовка до модульної контрольної роботи.

Питання що виносяться на модульну контрольну роботу:

1. Яким є основне завдання регресійного аналізу?
2. У чому полягають основні припущення класичного регресійного аналізу?
3. Якою є звичайна процедура класичного регресійного аналізу?
4. Як формулюється задача побудови регресійної моделі?
5. Які функціонали використовують для визначення параметрів регресійних моделей? У чому полягають переваги й недоліки різних типів таких функціоналів?
6. Якими є основні типи функцій, що використовуються для побудови однофакторних регресійних моделей?
7. Які моделі називають лінійними? Що називають порядком регресійної моделі?
8. Чому регресійні моделі не рекомендують використовувати поза межами тієї області значень вихідних параметрів, для якої вони побудовані?
9. Для заданого набору даних побудувати однофакторну лінійну регресійну модель і перевірити її адекватність.
10. У яких випадках нелінійні однофакторні моделі можна звести до лінійних? Навести приклади відповідних перетворень.
11. Для заданого набору даних побудуйте однофакторну нелінійну регресійну модель і перевірте її адекватність.
12. Як використовують критерій Фішера для перевірки адекватності регресійних моделей?
13. Як визначають довірчі інтервали для коефіцієнтів однофакторних регресійних моделей?
14. Яким є загальний вигляд поліноміальної регресійної моделі?
15. Яким є загальний алгоритм визначення порядку і параметрів поліноміальних регресійних моделей?
16. Для заданого набору даних побудуйте поліноміальну регресійну модель і перевірте її адекватність.
17. У яких випадках використовують регресійні моделі у вигляді тригонометричних поліномів? Яким є загальний алгоритм побудови таких моделей?
18. Для заданого набору даних побудуйте регресійну модель у вигляді тригонометричного поліному і перевірте її адекватність.
19. Якими є загальні алгоритми побудови однофакторних регресійних моделей у вигляді модифікованої показникової функції, кривої Гомперця та логістичної кривої?
20. Яким є загальний алгоритм побудови багатофакторної лінійної регресійної моделі?
21. Для заданого набору даних побудуйте багатофакторну лінійну регресійну модель і перевірте її адекватність.
22. Що називають мультиколінеарністю даних? Наведіть приклади.

23. Для чого застосовують алгоритми зміщеного оцінювання параметрів багатofакторних лінійних регресійних моделей? Наведіть приклади.
24. За якими властивостями перевіряють адекватність регресійних моделей? Якими є основні критерії адекватності?

ПРАКТИЧНЕ ЗАНЯТТЯ № 12

МОДУЛЬНА КОНТРОЛЬНА РОБОТА

Література:

1. Літнарівич Р.М. Побудова і дослідження математичної моделі за джерелами експериментальних даних методами регресійного аналізу. Навчальний посібник, МЕНУ, Рівне, 2011. 140 с.
2. Барковський В.В., Барковська Н.В., Лопатін О.К. Теорія ймовірностей та математична статистика : навч. посібник. 5-те вид. Київ : Центр учбової літератури, 2010. 424 с.
3. Блюмин С.Л., Суханов В.Ф., Чеботарьов С.В. Економічний факторний аналіз: монограф.. Липецьк: ЛЕГІ, 2004, 148 с.
4. Рогальський Ф.Б., Курилович Я.Є., Цокурєнка А. А. Математичні методи аналізу економічних систем. Книга 1. Київ: Наукова думка, 2001, 435 с.
5. Рогальський Ф.Б., Цокурєнка А.А. Математичні методи аналізу економічних систем. Книга 2. Київ: Наукова думка, 2001, 423 с.

Питання що виносяться на модульну контрольну роботу:

1. Яким є основне завдання регресійного аналізу?
2. У чому полягають основні припущення класичного регресійного аналізу?
3. Якою є звичайна процедура класичного регресійного аналізу?
4. Як формулюється задача побудови регресійної моделі?
5. Які функціонали використовують для визначення параметрів регресійних моделей? У чому полягають переваги й недоліки різних типів таких функціоналів?
6. Якими є основні типи функцій, що використовуються для побудови однофакторних регресійних моделей?
7. Які моделі називають лінійними? Що називають порядком регресійної моделі?
8. Чому регресійні моделі не рекомендують використовувати поза межами тієї області значень вихідних параметрів, для якої вони побудовані?
9. Для заданого набору даних побудувати однофакторну лінійну регресійну модель і перевірити її адекватність.
10. У яких випадках нелінійні однофакторні моделі можна звести до лінійних? Навести приклади відповідних перетворень.
11. Для заданого набору даних побудуйте однофакторну нелінійну регресійну модель і перевірте її адекватність.
12. Як використовують критерій Фішера для перевірки адекватності регресійних моделей?
13. Як визначають довірчі інтервали для коефіцієнтів однофакторних регресійних моделей?

14. Яким є загальний вигляд поліноміальної регресійної моделі?
15. Яким є загальний алгоритм визначення порядку і параметрів поліноміальних регресійних моделей?
16. Для заданого набору даних побудуйте поліноміальну регресійну модель і перевірте її адекватність.
17. У яких випадках використовують регресійні моделі у вигляді тригонометричних поліномів? Яким є загальний алгоритм побудови таких моделей?
18. Для заданого набору даних побудуйте регресійну модель у вигляді тригонометричного поліному і перевірте її адекватність.
19. Якими є загальні алгоритми побудови однофакторних регресійних моделей у вигляді модифікованої показникової функції, кривої Гомперця та логістичної кривої?
20. Яким є загальний алгоритм побудови багатофакторної лінійної регресійної моделі?
21. Для заданого набору даних побудуйте багатофакторну лінійну регресійну модель і перевірте її адекватність.
22. Що називають мультиколінеарністю даних? Наведіть приклади.
23. Для чого застосовують алгоритми зміщеного оцінювання параметрів багатофакторних лінійних регресійних моделей? Наведіть приклади.
24. За якими властивостями перевіряють адекватність регресійних моделей? Якими є основні критерії адекватності?

Навчальне видання

**МЕТОДИЧНІ ВКАЗІВКИ
ДО ПРАКТИЧНИХ ЗАНЯТЬ З НАВЧАЛЬНОЇ ДИСЦИПЛІНИ
«МЕТОДИ ОБРОБКИ КІЛЬКІСНИХ ПОКАЗНИКІВ ПОЖЕЖНОЇ
БЕЗПЕКИ»**

**для здобувачів третього (освітньо-наукового) рівня вищої освіти
галузь знань 26 «Цивільна безпека», спеціальність 261 «Пожежна
безпека»**

Підписано до друку __. __. __. Формат __ х __ / __.

Умовн.-друк. арк. __.

Вид. № __ / __.

Сектор редакційно-видавничої діяльності
Національного університету цивільного захисту України
61023 м. Харків, вул. Чернишевська, 94.
www.nuczu.edu.ua