

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
ХАРКІВСЬКИЙ НАЦІОНАЛЬНИЙ ЕКОНОМІЧНИЙ УНІВЕРСИТЕТ
ІМЕНІ СЕМЕНА КУЗНЕЦЯ

ЗАТВЕРДЖЕНО

на засіданні кафедри
статистики і економічного прогнозування
Протокол № 2 від 02.09.2024 р.



ПОГОДЖЕНО

Проректор з навчально-методичної роботи

Каріна НЕМАШКАЛО

ОСНОВИ АНАЛІЗУ ДАНИХ
робоча програма навчальної дисципліни (РПНД)

Галузь знань
Спеціальність
Освітній рівень
Освітня програма

12 Інформаційні технології
122 Комп'ютерні науки
другий (магістерський)
Комп'ютерні науки

Статус дисципліни

Мова викладання, навчання та оцінювання

вибіркова

українська

Розробники

д.е.н., професор
к.е.н., доцент
д.е.н., професор
викладач

підписано КЕП

Олена РАСВНСВА
Ольга БРОВКО
Костянтин СТРИЖИЧЕНКО
Мар'яна СЕМКІВ

Завідувач кафедри статистики і
економічного прогнозування

підписано КЕП

Олена РАСВНСВА

Гарант програми

підписано КЕП

Сергій МІНУХІН

Харків
2024

ВСТУП

Аналітика даних - це наука, що вивчає важливу інформацію та масиви великої кількості даних. Вона поєднує в собі статистичні методи аналізу даних із застосуванням ряду ІТ-технологій. Дисципліна пропонує широкий підхід до аналізу даних та методів пошуку даних та їх застосування на практиці. Вона об'єднує найсучасніші дослідження та практику в суміжних сферах та надає студентам необхідні знання та вміння ініціювати та вести аналіз даних, що може перетворити дані компанії в комерційно цінну інформацію.

Вивчення навчальної дисципліни «Основи аналізу даних» належить до циклу вибірових дисциплін освітнього ступеня «магістр» для здобувачів спеціальності 122 «Комп'ютерні науки» освітньо-професійної програми «Комп'ютерні науки» та спрямована на формування у здобувачів базових компетентностей щодо змоги аналізувати, об'єктивно оцінювати явища та процеси, встановлювати залежність результативного показника від комплексу факторів, виявляти тенденції зміни досліджуваних явищ і прогнозувати їх розвиток та інтерпретації отриманих результатів дослідження тощо.

Мета навчальної дисципліни: формування теоретичних знань, прикладних вмінь та навичок щодо організації статистичних спостережень, методів статистичного аналізу та прогнозування соціально-економічних явищ та процесів.

Основними завданнями вивчення дисципліни «Основи аналізу даних» є:

- визначення принципів проведення спостереження;
- опанування навичок обробки даних;
- вивчення методів побудови індексів та їх дослідження;
- формування вибірки та її аналіз з позиції репрезентативності;
- побудова індикаторів дослідження даних та їх аналіз.

Об'єктом навчальної дисципліни є соціально-економічні процеси та складові соціальних, культурних, економічних, інформаційних та бізнесових систем.

Предметом навчальної дисципліни є фундаментальні статистичні методи дослідження даних.

Результати навчання та компетентності, які формує навчальна дисципліна визначено в табл. 1.

Таблиця 1

Результати навчання та компетентності, які формує навчальна дисципліна

Результати навчання	Компетентності, яким повинен оволодіти здобувач вищої освіти
РН1. Мати спеціалізовані концептуальні знання, що включають сучасні наукові здобутки у сфері комп'ютерних наук і є основою для оригінального мислення та проведення досліджень, критичне осмислення проблем у сфері комп'ютерних наук та на межі галузей знань.	СК02. Здатність формалізувати предметну область певного проекту у вигляді відповідної інформаційної моделі.

РН7. Розробляти та застосовувати математичні методи для аналізу інформаційних моделей.	ЗК02. Здатність застосовувати знання у практичних ситуаціях
	СК03. Здатність використовувати математичні методи для аналізу формалізованих моделей предметної області.
РН8. Розробляти математичні моделі та методи аналізу даних (включно з великими).	СК03. Здатність використовувати математичні методи для аналізу формалізованих моделей предметної області.
РН9. Розробляти алгоритмічне та програмне забезпечення для аналізу даних (включно з великими).	ЗК02. Здатність застосовувати знання у практичних ситуаціях
	СК05. Здатність розробляти, описувати, аналізувати та оптимізувати архітектурні рішення інформаційних та комп'ютерних систем різного призначення.
РН19. Аналізувати сучасний стан і світові тенденції розвитку комп'ютерних наук та інформаційних технологій.	СК08. Здатність розробляти і реалізовувати проекти зі створення програмного забезпечення, у тому числі в непередбачуваних умовах, за нечітких вимог та необхідності застосовувати нові стратегічні підходи, використовувати програмні інструменти для організації командної роботи над проектом.
	СК12. Здатність розробляти, застосовувати та інтегрувати технології оброблення та аналізу даних в надпродуктивних системах та хмарних платформах для забезпечення ефективного використання обчислювальних ресурсів комп'ютерних систем.

ПРОГРАМА НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

Зміст навчальної дисципліни

Змістовий модуль 1. Методологічні основи аналізу даних

Тема 1. Природа спостереження за соціо-економічними об'єктами

1.1. Поняття статистичного спостереження. Формування інформаційної бази статистичного спостереження.

1.2. Метод аналітичних групувань.

1.3. Надання статистичних даних: таблиці, графіки, карти.

Тема 2. . Вибірка: її побудова та перевірка її репрезентативності

2.1. Сутність вибіркового спостереження, причини і умови його застосування.

2.2. Теоретичні основи вибіркового спостереження. Види і способи відбору одиниць у вибірку сукупність, що забезпечують репрезентативність вибірки.

2.3. Класифікація помилок вибірки. Порядок визначення помилок вибірки (репрезентативності) при різних способах відбору

2.4. Способи поширення результатів вибіркового спостереження на генеральну сукупність

2.5. Визначення необхідної чисельності вибірки.

Тема 3. Групування даних та дослідження груп

3.1. Групування, як основа науки про дані

3.2. Методи візуалізації згрупованих даних

Змістовий модуль 2. Методи дослідження та аналізу складових соціально-економічних систем

Тема 4. Індикатори дослідження даних

4.1. Сутність індикаторів дослідження.

4.2. Види та форми індикаторів (показників) при аналізі даних..

4.3. Способи та методи розрахунку індикаторів дослідження.

Тема 5. Світові індекси та їх дослідження

5.1. Індекси і їх класифікація

5.2. Загальні індекси кількісних показників.

5.3. Загальні індекси якісних показників.

5.4. Двохфакторна система взаємопов'язаних індексів.

5.5. Індексний метод аналізу динаміки середнього рівня.

5.6. Ланцюгові і базисні індекси.

Перелік лабораторних занять за навчальною дисципліною наведено в табл.

2

Таблиця 2

Перелік лабораторних занять

Назва теми та/або завдання	Зміст
Тема 1.	Лабораторна робота за темою 1. Формування репрезентативної вибірки.
Тема 2.	Лабораторне завдання за темою 2. Засоби візуалізації статистичних даних
Тема 3.	Лабораторна робота за темою 3. Групування даних та дослідження груп.
Тема 4.	Лабораторна робота за темою 4. Статистичні засоби дослідження динаміки явищ та процесів.
Тема 5.	Лабораторна робота за темою 5. Індикатори дослідження світових тенденцій.

Перелік самостійної роботи за навчальною дисципліною наведено в табл.

3.

Таблиця 3

Перелік самостійної роботи

Назва теми та/або завдання	Зміст
Тема 1-5	Вивчення лекційного матеріалу
Тема 1-2	Написання есе Вирішення ситуаційного (кейсового) завдання
Тема 1-5	Виконання лабораторних робіт

Кількість годин лекційних, лабораторних занять та годин самостійної роботи наведено в робочому плані (технологічній карті) з навчальної дисципліни.

МЕТОДИ НАВЧАННЯ

У процесі викладання навчальної дисципліни для набуття визначених результатів навчання, активізації освітнього процесу передбачено застосування таких методів навчання, як:

групова робота (Тема 1-2), проблемні лекції (Тема 1), створення ситуацій

пізнавальної новизни (теми 4-5).

Наочні (демонстрація (Тема 1-5)).

Практичні (лабораторна робота (Тема 2-5), есе (Тема 1), тестові завдання (Тема 1-5), тощо).

ФОРМИ ТА МЕТОДИ ОЦІНЮВАННЯ

Університет використовує 100 бальну накопичувальну систему оцінювання результатів навчання здобувачів вищої освіти.

Поточний контроль навчальної дисципліни здійснюється під час проведення лекційних, лабораторних занять і має на меті перевірку рівня підготовленості здобувача вищої освіти до виконання конкретної роботи і оцінюється сумою набраних балів:

— **для дисциплін з формою семестрового контролю екзамен (іспит):** максимальна сума – 60 балів; мінімальна сума, що дозволяє здобувачу вищої освіти скласти екзамен (іспит) – 35 балів

Підсумковий контроль включає семестровий контроль та атестацію здобувача вищої освіти.

Семестровий контроль проводиться у формах семестрового екзамену (іспиту). Складання семестрового екзамену (іспиту) здійснюється під час екзаменаційної сесії.

Максимальна сума балів, яку може отримати здобувач вищої освіти під час екзамену (іспиту) – 40 балів. Мінімальна сума, за якою екзамен (іспит) вважається складеним – 25 балів.

Підсумкова оцінка за навчальною дисципліною визначається:

— **для дисциплін з формою семестрового контролю екзамен (іспит) –** сумуванням балів за поточний та підсумковий контроль.

Під час викладання навчальної дисципліни використовуються наступні контрольні заходи:

Поточний контроль: Лабораторні роботи (25 балів), есе у вигляді презентації (10 балів), тестовий контроль (15 балів), письмові контрольні роботи (10 балів).

Семестровий контроль: Екзамен (40 балів)

Більш детальну інформацію щодо системи оцінювання наведено в робочому плані (технологічній карті) з навчальної дисципліни.

Приклад екзаменаційного білета

Харківський національний економічний університет імені Семена Кузнеця

Другий (магістерський) рівень вищої освіти

Спеціальність 122 «Комп'ютерні науки»

Освітньо-професійна програма «Комп'ютерні науки».

Навчальна дисципліна "Основи аналізу даних"

ЕКЗАМЕНАЦІЙНИЙ БІЛЕТ № 1

Стереотипне завдання (тести). (20 балів)

1	Показники, що характеризують обсяги, розміри соціально-економічних явищ, виражаються в одиницях вимірювання: а) натуральні, трудові та вартісні; б) коефіцієнти, відсотки, дюйми; в) всі відповіді вірні.
2	Перепис населення відноситься до спостереження: а) вибіркового; б) монографічного; в) суцільного; г) основного масиву.
3	Маркетингове дослідження адаптації нової марки товару на ринку відноситься до спостереження: а) поточне; б) одночасне; в) періодичне.
4	Індекс сезонності - це: а) процентне відношення окремих рівнів до середнього рівня даного ряду динаміки; б) відношення окремих рівнів до загальної суми рівнів ряду динаміки; в) ковзна середня окремих рівнів ряду.
5	Критерієм однорідності сукупності вважається: а) децильний коефіцієнт диференціації; б) квартильний коефіцієнт варіації; в) середньоквадратичний коефіцієнт варіації.
6	Можливе відхилення показників вибіркової сукупності від показників генеральної сукупності вимірюють: а) помилкою вибірки; б) дисперсією; в) середнім квадратичним відхиленням.
7	Статистична сукупність -це: а) явища і процеси суспільного життя; б) множина показників, що досліджуються статистикою; в) множина явищ, однорідних хоча б за однією будь-якою ознакою.
8	При обстеженні населення з приводу сумарних доходів сімей були опитані мешканці кожної четвертої квартири. Кожна четверта квартира була чоритьохкімнатною. Внаслідок цього результати обстеження сімей були невірними. Який вид помилки виникнув при цьому обстеженні? а) випадкова помилка репрезентативності; б) систематична помилка реєстрації; в) випадкова помилка репрезентативності; г) систематична помилка реєстрації.
9	Інвентаризація обладнання відноситься до спостереження: а) документальне спостереження; б) безпосереднє; в) опитування; г) експертні оцінки.
10	Дослідження якості продукції на підприємстві відноситься до спостереження: а) вибіркового; б) монографічного; в) суцільного;

	г) основного масиву.
11	Яким способом проводиться спостереження: а) прямим; б) задокументованим; в) опитування; г) експертні оцінки.
12	Структурне групування це а) тип угруповання, що дозволяє розділити групу на групи, характеризуючи її структуру за мінливою ознакою; б) поділ досліджуваного населення на соціально-економічні типи або класи; в) різновид групування, що дозволяє виявити наявність і спрямованість зв'язку ознак.
13	Проведено обстеження: 1) кожного 10-го сільськогосподарського підприємства із 150, які проводять зрошення земель за рахунок власних коштів з метою вивчення ефективності використання зрошувальних площ; 2) агрофірми "Коблево" з метою вивчення резервів підвищення ефективності зрошення саме у цьому господарстві. Які з обстежень є вибірковими? а) перше; б) друге; в) і перше, і друге; г) не має правильної відповіді.
14	Як відіб'ється на похибці вибірки збільшення дисперсії у 4 рази? а) похибка вибірки збільшиться у 4 рази; б) похибка вибірки зменшиться у 2 рази; в) похибка вибірки не зміниться; г) похибка вибірки збільшиться у 2 рази.
15	Статистичне спостереження – це а) планомірний науково організований збір даних про масові суспільні процеси та явища за допомогою реєстрації їх суттєвих ознак; б) реєстрація суспільних явищ і процесів в спеціальних формах статистичного спостереження; в) збір матеріалів спостереження та їх реєстрація.
16	Групування це: а) типологія соціально-економічних явищ; б) розподіл населення на групи істотних ознак; в) подання, обробка та підрахунок групових і загальних підсумків, що характеризують явище г) класифікація соціально-економічних явищ і процесів.
17	Яка із наведених формул зведених індексів є індексом фізичного обсягу експорту продукції: $\frac{\sum q_1 p_0}{\sum q_0 p_0}; \quad \frac{\sum q_1 p_1}{\sum q_1 p_0}; \quad \frac{\sum q_1 p_1}{\sum q_0 p_0}; \quad \frac{\sum q_0 p_0}{\sum q_1 p_0}.$ а) $\frac{\sum q_1 p_0}{\sum q_0 p_0}$ б) $\frac{\sum q_1 p_1}{\sum q_1 p_0}$ в) $\frac{\sum q_1 p_1}{\sum q_0 p_0}$ г) $\frac{\sum q_0 p_0}{\sum q_1 p_0}$
18	Встановлення дати початку та закінчення збору даних – це а) об'єктивний час спостереження; б) суб'єктивний час спостереження; в) критичний момент спостереження.
19	Медіана в ряду розподілу - це : а) найбільша частота (або значення ознаки); б) значення ознаки , що зустрічається найчастіше; в) значення ознаки , подумки ділить ряд розподілу на дві рівні частини.

20	Загальний індекс цін дорівнює $I_p = 0,92$. Як змінилась вартість продаж за рахунок цін? Відповіді: 1) збільшилась на 8%; 2) зменшилась на 8%; 3) змінилась на 0,92%. а) збільшилась на 8%; б) зменшилась на 8%; в) змінилась на 0,92%.
----	--

Діагностичне завдання 1 (розрахунковий тест). (3 бали)

Частка високоліквідних активів у сумі поточних активів становить 18%. Знайти дисперсію частки високоліквідних активів. Поясніть проведення розрахунків.

Діагностичне завдання 2 (есе). (7 балів)

Зробіть висновок щодо локалізації та концентрації зайнятих на підприємствах за вказаними видами економічної діяльності: коефіцієнт локалізації на підприємствах промисловості складає 3,4; на підприємствах сільського господарства – 5,7; на підприємствах торгівлі – 7,8. Коефіцієнт концентрації дорівнює 85%. Відповідь поясніть.

Евристичне завдання (розрахункове). (10 балів)

Маємо дані про обсяг товарообігу регіону, млн. грн.:

2017 рік	2018 рік	2019 рік	2020 рік	2021 рік	2022 рік	2023 рік
135	120	110	130	145	156	167

Провести аналіз даного динамічного ряду базисним та ланцюговим методами. Одержані показники подати у табличній формі. Розрахувати середньорічний абсолютний приріст, середньорічні темпи зростання та середньорічні темпи приросту товарообігу за 2017-2023 рр. Пояснити проведення та результати розрахунків. Зробити висновки за отриманими показниками.

Затверджено на засіданні кафедри статистики і економічного прогнозування протокол № _____ від «__» _____ 20__ р.

Екзаменатор

к.е.н., доц. Бровко О.І.

Зав. кафедрою

д.е.н., проф. Раєвська О. В.

Критерії оцінювання

Підсумкові бали за екзамен складаються із суми балів за виконання всіх завдань, що округлені до цілого числа за правилами математики.

Алгоритм вирішення кожного завдання включає окремі етапи, які відрізняються за складністю, трудомісткістю та значенням для розв'язання завдання. Тому окремі завдання та етапи їх розв'язання оцінюються відокремлено один від одного таким чином:

стереотипне завдання в формі закритих тестів – 20 тестів, максимальна оцінка за тест- 1 бал (максимальна кількість балів – 20);

діагностичне завдання 1 (розрахунковий тест) – максимальна оцінка 3 бали:

1 бал – підбір формули розрахунку;

1 бал – здійснення розрахункових дій;

1 бал – формулювання відповіді на поставлене запитання.

діагностичне завдання 2 (есе) – максимальна оцінка 7 балів:

1 бал – розуміння сутності завдання;

2 бали – підбір формул розрахунку або алгоритму виконання завдання;

2 бали – здійснення розрахунку або формулювання алгоритму дій по виконанню завдання;

2 бали - наявність аргументованих висновків за результатами аналітичної роботи;

евристичне завдання (розрахункове) – максимальна оцінка 10 балів:

2 бали – розуміння сутності завдання;

2 бали – підбір формул розрахунку, що відповідають сутності поставлених питань;

2 бали – розрахунок показників за окремими формулами стосовно теми дослідження;

2 бали – розрахунок показників в повному обсязі відповідно до поставлених питань у завданні;

2 бали – надання аргументованих висновків за результатами розрахункової роботи.

РЕКОМЕНДОВАНА ЛІТЕРАТУРА

Основна

1. Rayevnyeva O. Statistics [Electronic resource] : textbook / O. Rayevnyeva, I. Aksonova, O. Brovko [et al.]; Simon Kuznets Kharkiv national university of economics. - E-text data (3,53 МБ). - Kharkiv : S. Kuznets KhNUE, 2020. - 376 p. : il. - The title screen. - Referenc.: p. 356-362. <http://repository.hneu.edu.ua/handle/123456789/25678>

2. Галушак М. П., Галушак О. Я., Кужда Т. І. Прогнозування соціально-економічних процесів: навчальний посібник для економічних спеціальностей. – Тернопіль: ФОП Паляниця, 2021. – 160 с.

3. Гороховатський В. О. Методи інтелектуального аналізу та оброблення даних : навч. посіб. / В. О. Гороховатський, І. С. Творошенко ; М-во освіти і науки України, Харків. нац. ун-т радіоелектроніки. – Харків : ХНУРЕ, 2021. – 92 с.

4. Литвин В.В., Пасічник В.В. Аналіз даних та знань: навчальний посібник. К.: Магнолія 2006, 2023. - 276 с.

5. Муляр В. П. Візуалізація даних та інфографіка. Харків: ФОП Панов А. М. 2020. 200 с.

6. Статистичний аналіз даних : навчальний посібник / Т. М. Паянок, Т. М. Задорожня. – Ірпінь : Університет державної фіскальної служби України, 2020. – 312 с.

7. Чекотовський Е. В. Статистичні методи на основі Microsoft Excel 2016: навч. посібник. Київ: Знання, 2018. 408 с.

Додаткова

8. Rayevnyeva O. The Impact of a Sensitivity of Economic Activities on the Economic Behaviour of Enterprise / O. Rayevnyeva, S. Filip, I. Aksonova and other // Economics of Development. - 2022. - Vol. 21. - No. 3. - P. 27-39. <http://repository.hneu.edu.ua/handle/123456789/29607>

9. Аксьонова І. В. Статистичний моніторинг для розробки ключових показників ефективності підприємств та організацій / І. В. Аксьонова, Л. В. Ричкова // Актуальні аспекти сучасної статистичної науки і практики: матеріали Міжнародної науково-практичної конференції пам'яті проф. А.З. Підгорного, 31 травня 2023 р. : тези допов. – Одеса: Одеський національний економічний університет, 2023. – С. 20-22. <http://repository.hneu.edu.ua/handle/123456789/30968>
10. Аналіз даних та багатовимірна статистика : робоча програма навчальної дисципліни для здобувачів вищої освіти спеціальності 076 «Підприємництво та торгівля» освітньо-наукової програми «Підприємництво, торгівля та біржова діяльність» третього (освітньо-наукового) рівня: [Електронне видання] / укл. І. В. Аксьонова. – Харків : ХНЕУ ім. С. Кузнеця, 2024. – 8 с. (Укр. мов.) <http://repository.hneu.edu.ua/handle/123456789/32313>
11. Горват А.А., Молнар О.О., Мінькович В.В. Обробка, візуалізація та аналіз експериментальних даних з використанням пакету Origin: Навчальний посібник. Ужгород: Видавництво УЖНУ “Говерла”, 2020. – 64 с.:
12. Доценко С. І. Організація та системи керування базами даних: Навч. посібник. – Харків: УкрДУЗТ, 2023. – 117 с.
13. Корват О. В. Перспективи використання Microsoft Power BI у статистичному аналізі даних / О. В. Корват. // Розвиток бухгалтерського обліку, оподаткування і контролю в умовах інтеграційних процесів : матеріали Всеукр-ої наук.-практ. інтернет-конф., 22-23 жовт. 2020 р. – Херсон : ДВНЗ «ХДАУ», 2020. – С. 307-308 http://repository.hneu.edu.ua/bitstream/123456789/24849/1/%D0%9A%D0%BE%D1%80%D0%B2%D0%B0%D1%82_%D0%A5%D0%94%D0%90%D0%A3_2020.pdf
14. Прогнозування соціально-економічних процесів : конспект лекцій / уклад.: О. В. Шебаніна та ін. Миколаїв : МНАУ, 2022. – 95 с.
15. Статистичні методи та моделі оцінювання й прогнозування поведінки соціально-економічних систем в умовах інформаційної економіки : звіт кафедри статистики і економічного прогнозування за 2019-2020 навч. рік / викон. Раєвнева О. В., Аксьонова І. В., Бровко О. І. та ін. – Х.: ХНЕУ ім. С. Кузнеця, 2020. – 107 с. <http://repository.hneu.edu.ua/handle/123456789/23792>
16. Управління бізнес-процесами підприємства. Комплексний тренінг: навч. посіб. / П.Г. Банщиков, В.М. Гордієнко, О.О. Кизенко, Г.С. Скитьова. – К.: КНЕУ, 2018. – 283 с.

Інформаційні ресурси

17. Газін А. Візуалізація даних як навичка // А. Газін [Електронний ресурс]. – Режим доступу: https://aiukraine.com/wp-content/uploads/2017/10/1_4-Gazin.pdf
18. Офіційний сайт Всесвітнього Банку [Електронний ресурс]. – Режим доступу: <http://www.worldbank.org/>
19. Офіційний сайт Європейського центрального банку [Електронний ресурс]. – Режим доступу: <https://www.ecb.europa.eu/stats/html/index.en.html>

20. Офіційний сайт департаменту статистики Організації Об'єднаних Націй [Електронний ресурс]. – Режим доступу: <http://unstats.un.org/unsd/default.htm>

21. Сторінка курсу на платформі Moodle (персональна навчальна система).
- Режим доступу : <https://pns.hneu.edu.ua/course/view.php?id=4705>