

Київський університет імені Бориса Грінченка
Факультет інформаційних технологій та управління
Кафедра математики і фізики



«ЗАТВЕРДЖУЮ»
Проректор з науково-методичної та
навчальної роботи

Олексій ЖИЛЬЦОВ

2023р.

РОБОЧА ПРОГРАМА НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

АНАЛІТИКА ДАНИХ

для студентів

Спеціальності

111 Математика

Освітньої програми

111.00.02 Математичне моделювання

Освітнього рівня

другого (магістерського)

Київ – 2023

КИЇВСЬКИЙ УНІВЕРСИТЕТ ІМЕНІ БОРИСА ГРІНЧЕНКА
Ідентифікаційне посвідчення
Начальник відділу
моніторингу якості освіти

Протокол № 02/12/23
Жильцов
(підпис) (прізвище, ім'я, по-батькові)

23

Розробник:

Зінченко Надія Мусіївна, доктор фізико-математичних наук, професор кафедри математики і фізики Факультету інформаційних технологій та математики Київського університету імені Бориса Грінченка

Викладач:

Зінченко Надія Мусіївна, доктор фізико-математичних наук, професор кафедри математики і фізики Факультету інформаційних технологій та математики Київського університету імені Бориса Грінченка

Робочу програму розглянуто і затверджено на засіданні кафедри математики і фізики

Протокол від 23 серпня 2023 р. № 8

Завідувач кафедри *Сем* Світлана Семеняка

Робочу програму погоджено з гарантом освітньої програми (керівником освітньої програми 111.00.02 Математичне моделювання
13.08. 2023 р.

Керівник освітньої програми *В. Прошкін* Володимир ПРОШКІН
(підпис)

Робочу програму перевірено

 . . 20*23* р.

Заступник директора/декана *С. Іванченко* Євген ІВАНІЧЕНКО

Пролонговано:

на 20__/20__ н.р. (підпис) ((ПІБ)), « » 20__ р., протокол №

на 20__/20__ н.р. (підпис) ((ПІБ)), « » 20__ р., протокол №

на 20__/20__ н.р. (підпис) ((ПІБ)), « » 20__ р., протокол №

на 20__/20__ н.р. (підпис) ((ПІБ)), « » 20__ р., протокол №

1. Опис навчальної дисципліни

| Найменування показників | Характеристика дисципліни за формами навчання |
|---|---|
| | денна |
| Вид дисципліни | вибіркова |
| Мова викладання, навчання та оцінювання | українська |
| Загальний обсяг кредитів / годин | 5/150 год. |
| Курс | 2 |
| Семестр | 3 |
| Кількість змістових модулів з розподілом: | 4 |
| Обсяг кредитів | 5 |
| Обсяг годин, в тому числі: | 150 год. |
| Аудиторні | 40 год. |
| Модульний контроль | 10 год. |
| Самостійна робота | 100 год. |
| Форма семестрового контролю | залік |

2. Мета та завдання навчальної дисципліни

Метою викладання дисципліни є підготовка студентів до ефективного застосування інструментів аналітики даних у фаховій практичній діяльності

Мета досягається через практичне оволодіння студентами навичками роботи з новітніми інформаційними технологіями у сфері аналітики даних.

Завдання:

- отримати знання, уміння і набути навички, необхідні для аналізу даних;
- засвоїти методiku і напрями використання сучасних інструментів для аналізу даних;
- набути уміння та навички ефективно використовувати Power BI для аналізу даних;
- навчитись використовувати набуті знання, уміння та навички для ілюстрації результатів дослідження.

Загальні компетентності:

ЗК-2: Критичне мислення. Здатність аналізувати, верифікувати, оцінювати повноту та достовірність інформації в ході професійної діяльності, за необхідності доповнювати й синтезувати відсутню інформацію.

ЗК-3: Креативність. Продукування нових ідей, творчий підхід до їх реалізації; здатність до новаторської діяльності.

ЗК-8: Когнітивна гнучкість. Здатність здобувати нові знання, уміння та інтегрувати їх з уже наявними; спроможність аналізувати явище, ситуацію, проблему, враховуючи різні параметри, фактори, причини; здатність адаптувати мислення для вирішення задач у змінених умовах чи нестандартних ситуаціях.

Фахові компетентності:

ФК-2: Дослідницькі навички. Здатність розуміти сутність проблеми, постановку задачі, обирати та використовувати відповідні методи й організаційні процедури для її вирішення (розв'язання), дослідницької чи інноваційної діяльності, критично оцінювати отримані результати, визначати перспективи подальшої розробки досліджуваної та дотичних тем.

ФК-8: Самоосвіта та підвищення кваліфікації. Здатність до самоосвіти та підвищення кваліфікації у сфері математики, дидактики, освітніх технологій на основі інноваційних підходів.

3. Результати навчання за дисципліною

За результатами вивчення дисципліни у студента буде сформовано такі загальні компетентності як: критичне мислення, креативність, когнітивна гнучкість та такі фахові

компетентності як: дослідницькі навички, самоосвіта та підвищення кваліфікації.

У результаті вивчення дисципліни студент повинен знати:

- основні етапи аналізу даних; класифікацію типів змінних; поняття кількісних, ординальних та номінальних даних; вибіркові характеристики за згрупованими даними;
- визначення головних характеристик скалярних змінних: центру групування значень, розсіювання значень, скошеності та гостровершинності розподілу; характеристики векторних спостережень; методи візуалізації даних; графічні експрес-методи аналізу даних;
- характеристики парного та множинного статистичного зв'язку кількісних даних: коефіцієнт кореляції, індекс кореляції, коефіцієнт детермінації, множинний коефіцієнт кореляції; частинні коефіцієнти кореляції; аналіз рангових кореляцій; парні рангові коефіцієнти кореляції; коефіцієнт конкордації; кореляційний аналіз номінальних даних; аналіз таблиць спряженості; коефіцієнти спряженості; інформаційна міра зв'язку; оцінювання та перевірка на значимість характеристик статистичного зв'язку;
- основні задачі регресійного аналізу; властивості оцінок характеристик моделі та основні допоміжні твердження; поняття довірчої області для параметрів моделі; значимість параметрів моделі; адекватність моделі; оцінювання при наявності лінійних обмежень гіпотези у лінійній регресії; частинна f -статистика; класичні припущення (мультиколінеарність, неоднорідність та корельованість похибок); поняття оптимальної регресійної моделі;
- моделі дисперсійного аналізу; однофакторний дисперсійний аналіз; багатофакторний дисперсійний аналіз; поняття впливів головних ефектів та взаємодій;
- поняття простору об'єктів, спостережень, ознак; основні постановки задач класифікації; дискримінантний аналіз; поняття навчаючої вибірки; дискримінантні функції; класифікацію з учителем у випадку нормальних спостережень; підстановочне розв'язуюче правило; кластерний аналіз; поняття «аналізу корзин»; дерева рішень;
- поняття тренду та стохастичної складової часового ряду, поліноміального тренду часового ряду, порядку поліноміального тренду періодичного тренду; спектрограми, періодограми; метод ковзного середнього; означення спектральної щільності,
- особливості планування і аналізу результатів вибірових обстежень.

Студент повинен вміти:

- класифікувати типи змінних; знаходити кількісні, ординальні та номінальні дані групувати дані; отримувати вибіркові характеристики за згрупованими даними;
- визначати головні характеристики скалярних змінних: центру групування значень, розсіювання значень, скошеності та гостровершинності розподілу; знаходити характеристики векторних спостережень; оцінювати параметри основних ймовірнісних розподілів; перевіряти основні гіпотези; знаходити і видаляти аномальні спостереження у скалярних та векторних вибірках; перевіряти стохастичність вибірки; володіти основними методами розв'язувального аналізу; візуалізувати дані; використовувати графічні експрес-методи аналізу даних;
- знаходити характеристики парного та множинного статистичного зв'язку кількісних даних: коефіцієнт кореляції, індекс кореляції, коефіцієнт детермінації, кореляційне відношення, множинний коефіцієнт кореляції; знаходити частинні коефіцієнти кореляції; аналізувати рангові кореляції; знаходити ранг, об'єднаний ранг; створювати таблиці рангів; знаходити парні рангові коефіцієнти кореляції, коефіцієнт конкордації; виконувати кореляційний аналіз номінальних даних, аналіз таблиць спряженості; знаходити коефіцієнти спряженості, інформаційну міру зв'язку; оцінювати та перевіряти на значимість характеристики статистичного зв'язку;
- розв'язувати основні задачі регресійного аналізу; знаходити властивості оцінок характеристик моделі; знаходити довірчі області для параметрів моделі, інтервали Бонфероні; перевіряти на значимість параметри моделі; перевіряти на адекватність моделі; оцінювати

модель при наявності лінійних обмежень; перевіряти гіпотези у лінійній регресії; знаходити частинну f-статистику; розв'язувати задачі у випадку порушення класичних припущень (мультиколінеарності, неоднорідності та корельованості похибок); здійснювати вибір оптимальної регресійної моделі;

- створювати моделі дисперсійного аналізу; виконувати однофакторний дисперсійний аналіз; знаходити контрасти та довірчі інтервали для них; виконувати багатофакторний дисперсійний аналіз; визначати впливи головних ефектів та взаємодій;

- створювати модель та знаходити її зв'язок з моделями регресійного та дисперсійного аналізу; виконувати двокроковий метод найменших квадратів;

- будувати простори об'єктів, спостережень, ознак; формулювати задачі класифікації; виконувати дискримінантний аналіз; будувати навчаючу вибірку; знаходити дискримінантні функції; будувати класифікацію з учителем у випадку нормальних спостережень; формулювати підстановочне розв'язуюче правило; виконувати кластерний аналіз.

- знаходити тренд та стохастичну складову часового ряду; оцінювати поліноміальний тренд часового ряду; визначати порядок поліноміального тренду; аналізувати часові ряди з періодичними трендами; будувати спектрограму, періодограму; застосовувати метод ковзного середнього; визначати спектральну щільність.

та досягти наступних програмних результатів навчання:

ПРН-3-2: Володіти основами математичних дисциплін і теорій, які вивчають моделі природничих, економічних і соціальних процесів.

ПРН-3-4: Демонструвати знання й розуміння зв'язку окремих розділів теоретичної та прикладної математики із економічними процесами і теоріями для побудови ефективних економічних моделей.

ПРН-У-1: Коректно проводити логічні міркування, грамотно вибудовувати доведення математичних фактів, використовуючи, в тому числі, класичні методи доведення (від супротивного, математичної індукції, конструктивний та ін.).

ПРН-У-3: Демонструвати вміння використовувати фундаментальні математичні закономірності при розв'язуванні теоретичних та прикладних математичних задач і проблем, які потребують, зокрема, інтеграції набутих знань, методів з різних розділів математики, в т.ч. багатокритеріальні задачі та задачі з неповними даними.

ПРН-У-13: Уміти формулювати математичну / педагогічну задачу, знаходити й аналізувати відповідності між поставленою задачею й існуючими моделями, аргументовано обирати оптимальні шляхи розв'язання, аналізувати й осмислювати отриманий розв'язок, представляти результати роботи й обґрунтовувати запропоновані рішення на сучасному науково-технічному й професійному рівні.

4. Структура навчальної дисципліни

| Назви змістових модулів, тем | Ус бог о | Розподіл годин між видами робіт | | |
|--|----------------|---------------------------------|----------|-----------|
| | | Аудиторна | | С.Р. |
| | | Л | П. | |
| Змістовий модуль 1. Попередня обробка даних | | | | |
| Предмет аналізу даних | 6 | 1 | - | 5 |
| Попередня обробка даних | 17 | 3 | 4 | 10 |
| <i>Модульний контроль</i> | 2 | | | |
| Разом за змістовим модулем 1 | | | | |
| | 25 | 4 | 4 | 15 |
| Змістовий модуль 2. Методи дослідження і моделювання взаємозв'язків між кількісними і якісними факторами | | | | |
| Застосування кореляційного і регресійного аналізу у природничих науках, при моделюванні економічних і соціальних процесів та в освітніх вимірюваннях | 14 | 2 | 2 | 10 |

| | | | | |
|--|------------|-----------|-----------|------------|
| Факторний аналіз | 14 | 2 | 2 | 10 |
| Дисперсійний аналіз | 14 | 2 | 2 | 10 |
| Коваріаційний аналіз | 14 | 2 | 2 | 10 |
| <i>Модульний контроль</i> | 4 | | | |
| <i>Разом за змістовим модулем 2</i> | 60 | 8 | 8 | 40 |
| Змістовий модуль 3. Задачі класифікації | | | | |
| Дискримінантний аналіз | 14 | 2 | 2 | 10 |
| Кластерний аналіз | 14 | 2 | 2 | 10 |
| <i>Модульний контроль</i> | 2 | | | |
| <i>Разом за змістовим модулем 3</i> | 30 | 4 | 4 | 20 |
| Змістовий модуль 4. Аналіз часових рядів. Теорія вибірових обстежень. | | | | |
| Аналіз часових рядів | 19 | 2 | 2 | 15 |
| Теорія вибірових обстежень | 14 | 2 | 2 | 10 |
| <i>Модульний контроль</i> | 2 | | | |
| <i>Разом за змістовим модулем 4</i> | 35 | 4 | 4 | 25 |
| Усього годин | 150 | 20 | 20 | 100 |

5. Програма навчальної дисципліни

Змістовий модуль 1. Попередня обробка даних .

Тема 1. Предмет та основні задачі аналізу даних.

Етапи аналізу даних. Розвідувальний і підтверджуючий аналіз. Класифікація типів змінних.

Кількісні і якісні, ординальні та номінальні дані та робота з ними. Групування даних.

Емпірична функція розподілу. Гістограма. Знаходження вибірових характеристик за згрупованими даними.

Тема 2. Попередня обробка даних .

Визначення головних вибірових характеристик скалярних змінних: положення (середнього, центру групування значень), розсіювання (вар'ювання) значень, скошеності та гостровершинності розподілу (коефіцієнтів асиметрії і ексцесу) і т.д. Характеристики векторних спостережень. Аналіз однієї та двох вибірок: оцінювання параметрів і перевірка гіпотез. Розвідувальний аналіз. Методи візуалізації даних. Графічні експрес-методи аналізу даних.

Змістовий модуль 2. Методи дослідження і моделювання взаємозв'язків між кількісними та якісними факторами .

Тема 3. Застосування кореляційного і регресійного аналізу у природничих науках, при моделюванні економічних і соціальних процесів та в освітніх вимірюваннях.

Інструменти і стратегії дослідження ознак, характеристики парного та множинного статистичного зв'язку кількісних даних, кореляційний аналіз номінальних даних, оцінювання та перевірка на значимість характеристик статистичного зв'язку. Специфіка застосування регресійного аналізу в різних галузях знань.

Тема 4. Факторний аналіз.

Сутність факторного аналізу. Зменшення розмірності простору ознак. Метод головних компонент.

Тема 5. Дисперсійний аналіз.

Дослідження впливу якісних факторів. Моделі дисперсійного аналізу. Однофакторний дисперсійний аналіз. Контрасти та довірчі інтервали для них. Багатофакторний дисперсійний аналіз. Визначення впливів головних ефектів та взаємодій.

Тема 6. Коваріаційний аналіз.

Вивчення впливу кількісних і якісних факторів. Основна модель та її зв'язок з моделями регресійного та дисперсійного аналізу. Двокроковий метод найменших квадратів.

Змістовий модуль 3. Задачі класифікації.

Тема 7. Дискримінантний аналіз.

Простори об'єктів, спостережень, ознак. Основні постановки задач класифікації.

Дискримінантний аналіз. Навчаюча вибірка. Дискримінантні функції. Класифікація з учителем у випадку нормальних спостережень. Підстановочне розв'язуюче правило. Пошук асоціативних правил.

Тема 8. Кластерний аналіз. Поняття кластеризації. Особливості задачі кластеризації. Поняття схожості та відстані між об'єктами у просторі атрибутів. Види відстаней між об'єктами.

Агломеративні та дивізимні алгоритми кластеризації. Дендрограма. алгоритм k середніх.

Змістовний модуль 4. Аналіз часових рядів. Теорія вибірових обстежень.

Тема 9. Аналіз часових рядів.

Тренд та стохастична складова часового ряду. Оцінювання поліноміального тренду часового ряду. Визначення порядку поліноміального тренду. Аналіз часових рядів з періодичними трендами. Спектрограма, періодограма. Метод ковзного середнього. Визначення спектральної щільності.

Тема 10. Теорія вибірових обстежень.

Специфіка скінченної генеральної сукупності. Планування і аналіз даних.

6. Контроль навчальних досягнень

6.1. Система оцінювання навчальних досягнень студентів

Розрахунок рейтингових балів за видами поточного (модульного) контролю

| № з/п | Вид діяльності студента | Макс. кількість балів за одиницю | Модуль 1 | | Модуль 2 | | Модуль 3 | | Модуль 4 | |
|-------|---|----------------------------------|---------------------------------|------------------------------|------------------------------|------------------------------|------------------------------|------------------------------|------------------------------|------------------------------|
| | | | Кільк. одиниць до розрахунку | Макс. кількість балів за вид | Кільк. одиниць до розрахунку | Макс. кількість балів за вид | Кільк. одиниць до розрахунку | Макс. кількість балів за вид | Кільк. одиниць до розрахунку | Макс. кількість балів за вид |
| 1 | Відвідування лекцій | 1 | 1 | 1 | 4 | 4 | 2 | 2 | 2 | 2 |
| 2 | Відвідування практичних занять | 11 | 3 | 3 | 4 | 4 | 2 | 2 | 2 | 2 |
| 3 | Виконання завдань для самостійної роботи | 5 | 1 | 5 | 1 | 5 | 1 | 5 | 1 | 5 |
| 4 | Робота на практичних (семінарських) заняттях | 10 | 3 | 30 | 3 | 30 | 2 | 20 | 2 | 20 |
| 5 | Виконання модульної контрольної роботи | 25 | 1 | 25 | 2 | 50 | 1 | 25 | 1 | 25 |
| 6 | Лабораторне заняття (допуск, виконання, захист) | 10 | - | - | - | - | - | - | - | - |
| 7 | ІНДЗ | 30 | - | - | - | - | - | - | - | - |
| | Макс. кількість балів за видами поточного контролю (МВ) | | | 64 | | 93 | | 54 | | 54 |
| | Розрахунок коефіцієнту | | $\frac{100}{265} = 0,377358491$ | | | | | | | |

6.2. Завдання для самостійної роботи та критерії її оцінювання.

Завдання для самостійної роботи передбачає самостійне виконання завдань в межах кожного змістового модуля.

| № з/п | Назва теми | Кількість годин | Кількість балів |
|-------|-------------------------|-----------------|-----------------|
| 1 | Попередня обробка даних | 25 | 5 |
| 2 | Факторний аналіз. | 25 | 5 |
| 3 | Кластерний аналіз | 25 | 5 |
| 4 | Вибіркові обстеження | 25 | 5 |
| | Разом | 100 | 20 |

Критерії оцінювання:

- 5 балів – правильно виконано 5 завдань,
- 4 балів – правильно виконано 4 завдання,
- 3 балів – правильно виконано 3 завдання,
- 2 бали – правильно виконано 2 завдання,
- 1 бали – правильно виконано 1 завдання.

6.3. **Форми проведення модульного контролю та критерії оцінювання**

Форма проведення модульного контролю – виконання тестів за комп'ютером. Критерії оцінювання: кожне правильно виконане завдання оцінюється у 5 балів.

6.4. **Форми проведення семестрового контролю та критерії оцінювання**

Підсумковий контроль проводиться у формі заліку. Бали просумовуються протягом вивчення дисципліни за всі види занять.

6.5. **Шкала відповідності оцінок**

| Рейтингова оцінка | Оцінка за стобальною шкалою | Значення оцінки |
|-------------------|-----------------------------|---|
| A | 90 – 100 балів | Відмінно – відмінний рівень знань (умінь) в межах обов'язкового матеріалу з можливими незначними недоліками |
| B | 82-89 балів | Дуже добре – достатньо високий рівень знань (умінь) в межах обов'язкового матеріалу без суттєвих (грубих) помилок |
| C | 75-81 балів | Добре – в цілому добрий рівень знань (умінь) з незначною кількістю помилок |
| D | 69-74 балів | Задовільно – посередній рівень знань (умінь) із значною кількістю недоліків, достатній для подальшого навчання або професійної діяльності |
| E | 60-68 балів | Достатньо – мінімально можливий допустимий рівень знань (умінь) |
| FX | 35-59 балів | Незадовільно з можливістю повторного складання – незадовільний рівень знань, з можливістю повторного перескладання за умови належного самостійного доопрацювання |
| F | 1-34 балів | Незадовільно з обов'язковим повторним вивченням курсу – досить низький рівень знань (умінь), що вимагає повторного вивчення дисципліни |

7. Рекомендовані джерела

Базова

1. М.В. Карташов. Теорія ймовірностей та математична статистика. К.: ВПЦ "Київський університет", 2007.
2. А.Я. Оленко. Комп'ютерна статистика. К.: ВПЦ "Київський університет", 2007.
3. В.М Турчин. Математична статистика в прикладах і задачах. - К., 1993.
4. В.М. Турчин. Теорія ймовірностей і математична статистика. Вид-во Дніпро. НУ, 2006..
5. Жлуктенко В. І., Наконечний С. І. Теорія ймовірностей і математична статистика: Навч.-метод. посібник: У 2 ч. – К.: КНЕУ, 2000. – Ч. І: Теорія ймовірностей. – 304
6. Жолдак М. І., Кузьміна Н. М., Берлінська С. Ю. Теорія ймовірностей і математична статистика. – К.: Вища школа, 1995. – 351 с
7. Сидорова А. В., Біленко Д. В., Буркіна Н. В. Бізнес-аналітика: навчально-методичний посібник. Вінниця: ДонНУ імені Василя Стуса. 2019

Допоміжна

1. Борковський В. В., Барковська Н. В., Лопатін О. К. Теорія ймовірностей і математична статистика. – К.: ЦУЛ, 2002. – 448 с. – (Математичні науки).
2. R. D'Agostino, V. Stephens. Goodness-of-fit Techniques, Marcel Dekker, NY, 1986
3. М.В. Карташов. Імовірність, процеси, статистика. К.: ВПЦ "Київський університет", 2007.
4. Бугір М. К. Теорія ймовірностей і математична статистика: Навч. посібник. – Тернопіль: Підручники і посібники, 1998. – 176 с.
5. Бугір М. К. Практикум з теорії ймовірностей та математичної статистики: Навч. посібник. – Тернопіль: УМДС, 1998.
6. Жлуктенко В. І., Наконечний С. І. Теорія ймовірностей з елементами математичної статистики. – К.: УМКВО, 1991.
7. Копич І. М. Елементи теорії ймовірностей і математичної статистики: Навч. посібник. – Л.: Коопосвіта, ЛКА, 1997.
8. Chris Chatfield . The Analysis of Time Series, an Introduction (вид. 5-те). Chapman & Hall. 1996.
9. Єріна А.М. Статистичне моделювання та прогнозування: Навч. Посібник / А.М. Єріна – К.: КНЕУ, 2014. – 340 с