



# Обробка даних методами дисперсійного аналізу

## Робоча програма навчальної дисципліни (Силабус)

• Реквізити навчальної дисципліни

Рівень вищої освіти	Другий (освітньо-науковий)
Галузь знань	11 Математика та статистика
Спеціальність	111 Математика
Освітня програма	Математика Mathematics
Статус дисципліни	Нормативна
Форма навчання	Очна (денна)
Рік підготовки, семестр	2 курс, осінній семестр
Обсяг дисципліни	4 кредити (120 годин), з них аудиторні – 36 годин лекційні, 18 годин практичні заняття, СРС – 66 годин
Семестровий контроль/ контрольні заходи	Іспит/МКР, РГР
Розклад занять	<a href="http://rozklad.kpi.ua/Schedules/ScheduleGroupSelection.aspx">http://rozklad.kpi.ua/Schedules/ScheduleGroupSelection.aspx</a>
Мова викладання	Українська
Інформація про керівника курсу / викладачів	Лектор: д. ф.-м. н., проф. Іванов Олександр Володимирович, e-mail: <a href="mailto:alexntuu@gmail.com">alexntuu@gmail.com</a> , тел. +38(067)9668421 Практичні / Семінарські: к.ф.-м.н. доцент Орловський Ігор Володимирович, e-mail: <a href="mailto:i.v.orlovsky@gmail.com">i.v.orlovsky@gmail.com</a> , тел. +38(098)4833292
Розміщення курсу	<a href="https://campus.kpi.ua">https://campus.kpi.ua</a>

• Програма навчальної дисципліни

1. Опис навчальної дисципліни, її мета, предмет вивчення та результати навчання

<b>Цілі дисципліни</b>	Дисперсійний аналіз (ANOVA) – це метод математичної статистики, спрямований на пошук залежностей у вибіркових даних шляхом дослідження значущості відмінностей у середніх значеннях. Мета навчальної дисципліни полягає в ознайомленні майбутніх фахівців-математиків з однофакторним, двофакторним та багатофакторним ANOVA, тобто дослідженням значущості розбіжності між середніми шляхом порівняння (аналізу) дисперсій. ANOVA відноситься до параметричних методів у статистиці, його слід застосовувати тільки тоді, коли розподіли спостережень є гауссівськими.
<b>Предмет навчальної дисципліни</b>	Предметом навчальної дисципліни є співвідношення систематичної (міжгрупової) дисперсії та випадкової (внутрішньогрупової) дисперсії у вибірках. У якості показника мінливості використовується сума квадратів відхилень значень параметрів від середнього. Співвідношення міжгрупової та внутрішньогрупової дисперсії має розподіл Фішера за умови незалежності та гауссовості вибірок і визначається за допомогою F-критерія Фішера.
<b>Компетентності</b>	ЗК1 Здатність учитися, здобувати нові знання, уміння, у тому числі в галузях, відмінних від математики ЗК2 Здатність використовувати у професійній діяльності знання з галузей математичних, природничих, соціально-гуманітарних та економічних наук ЗК3 Здатність вирішувати проблеми у професійній діяльності на основі абстрактного мислення, аналізу, синтезу та прогнозу ФК16 Володіння знаннями та здатність ініціювати й проводити наукові дослідження у спеціалізованій області математики ФК17 Спроможність займатись науковою та дослідницькою діяльністю, використовуючи при цьому новітні технології та інноваційні підходи у сфері математики ФК18 Здатність використовувати інформаційно-комунікаційні технології та спеціалізоване програмне забезпечення у навчальному процесі.
<b>Програмні результати навчання</b>	РН14 Знати головні результати та сфери застосувань основних математичних теорій, що використовуються при математичному моделюванні: фінансової та актуарної математики, методів математичної економіки та імітаційного моделювання, комп'ютерної статистики РН20 Знати теоретичні основи методів дисперсійного аналізу даних, знати функції, за допомогою яких ці методи реалізовані у спеціалізованому програмному забезпеченні, та вміти їх застосувати для аналізу даних у прикладних задачах

## **2. Пререквізити та постреквізити дисципліни (місце в структурно-логічній схемі навчання за відповідною освітньою програмою)**

### **Пререквізити.**

Нормативний компонент освітньої програми ЗО7: «Обробка даних методами дисперсійного аналізу» відноситься до циклу загальної підготовки здобувачів 2-го (освітньо-наукового) рівня вищої освіти і базується на знаннях дисциплін “Вступ до теорії ймовірностей”, “Теорія ймовірностей”, “Основи математичної статистики”, які викладаються на бакалаврському рівні і дисципліни ПОЗ “Прикладні моделі нелінійного дисперсійного аналізу”.

### **Постреквізити.**

Знання та уміння, отримані здобувачем під час вивчення цієї дисципліни, будуть потрібні їм при написанні тексту магістерської дисертації та подальшої роботи за фахом.

## **3. Зміст навчальної дисципліни**

Кредитний модуль включає наступні теми.

Розділ 1. Деякі результати з теорії множинної лінійної регресії.

Тема 1.1. Статистична незалежність оцінки найменших квадратів (LSE) параметрів регресії та залишкової суми квадратів (RSS).

Тема 1.2. Критерій Фішера (F-критерій) перевірки лінійних гіпотез щодо невідомого параметра множинної лінійної регресії.

Розділ 2. Однофакторний ANOVA.

Тема 2.1. Перевірка гіпотези рівності середніх двох нормальних сукупностей.

Тема 2.2. Перевірка гіпотези рівності середніх декількох нормальних сукупностей.

Розділ 3. Двофакторний ANOVA.

Тема 3.1. Класифікація даних за двома ознаками.

Тема 3.2. Загальна модель даних з двома ознаками та компоненти дисперсії.

Розділ 4. Багатофакторний ANOVA на прикладі трифакторного ANOVA.

Тема 4.1. Головні ефекти та взаємодії у трифакторному ANOVA.

Тема 4.2. Таблиця трифакторного ANOVA.

Розділ 5. Деякі питання планування експерименту в ANOVA.

Тема 5.1. Рандомізований блок-дизайн.

Тема 5.2. Схема латинського квадрата.

Розділ 6. Дворівневі факторні плани.

Тема 6.1. Повний факторний експеримент  $2^m$  (ПФЕ  $2^m$ ) та його репліки.

## **4. Навчальна література та ресурси.**

### **1. Базова література**

1. G.W. Oehlert. A First Course in Design and Analysis of Experiments, University of Minnesota, 2010, 679 p. [Електронний ресурс] – режим доступу до ресурсу: <http://users.stat.umn.edu/~gary/book/fcdae.pdf>
2. Analysis of Variance, HELM, Workbook 44, 2015, 62 p. [Електронний ресурс] – режим доступу до ресурсу: [http://nucinkis-lab.cc.ic.ac.uk/HELM/HELM\\_Workbooks\\_41-45/WB44-all.pdf](http://nucinkis-lab.cc.ic.ac.uk/HELM/HELM_Workbooks_41-45/WB44-all.pdf)

3. D.M. Lane. Analysis of Variance, HyperStat Online Statistics Textbook, Rice University, 2013, pp. 517-598. [Електронний ресурс] – режим доступу до ресурсу: [http://onlinestatbook.com/2/analysis\\_of\\_variance/anova.pdf](http://onlinestatbook.com/2/analysis_of_variance/anova.pdf)
4. B.G. Tabachnick, L.S. Fidell. Experimental Designs using ANOVA, California State University, Northridge, 2020, 770 p. [Електронний ресурс] – режим доступу до ресурсу: [http://www.researchgate.net/profile/Barbara-Tabachnick/publication/259465542\\_Experimental\\_Designs\\_Using\\_ANOVA/links/5e6bb05f92851c6ba70085db/Experimental-Designs-Using-ANOVA.pdf](http://www.researchgate.net/profile/Barbara-Tabachnick/publication/259465542_Experimental_Designs_Using_ANOVA/links/5e6bb05f92851c6ba70085db/Experimental-Designs-Using-ANOVA.pdf)
5. A.H. Kvanli, R.J. Pavur, C.S. Guynes, Introduction to business statistics (5<sup>th</sup> edn.), Cincinnati: South Western College Publishing, 2020, 930 p.

## 2. Допоміжна література

6. H. Scheffe. Analysis of variance. Wiley Classics Library Edition, 1999.
7. S.R. Rao. Linear statistical inference and its application, Willey, 2<sup>nd</sup> Edition, 1973, 552 p.
8. R. Balkin. Post hoc analysis for ANOVA, Texas A&M University – Commerce, 2008, 16 p. [Електронний ресурс] – режим доступу до ресурсу: [http://www.balkinresearchmethods.com/Balkin\\_Research\\_Methods/Research\\_Methods\\_and\\_Statistics\\_files/Post%20hoc%20analysis%20for%20ANOVA.pdf](http://www.balkinresearchmethods.com/Balkin_Research_Methods/Research_Methods_and_Statistics_files/Post%20hoc%20analysis%20for%20ANOVA.pdf)
9. J.L. Devore, K.N. Berk. Modern mathematical statistics with applications, 2<sup>nd</sup> Edition, Springer, 2012, 845 p.

## ● Навчальний контент

### 5. Методика опанування навчальної дисципліни (освітнього компонента)

#### Лекції

№ з/п	Назва теми лекції та перелік основних питань (перелік дидактичних матеріалів, посилання на літературу)
1.1.1	<b>Геометричний сенс оцінювання методом найменших квадратів параметрів множинної регресії.</b> Застосування теореми Кохрена до розкладу за теоремою Піфагора суми квадратів похибок спостережень у просторі $R^N$ . <i>Рекомендована література:</i> [2], с. 6–14; [5], с.385–406.
1.1.2	<b>Незалежність RSS від LSE.</b> Побудова надійних інтервалів для невідомих параметрів та дисперсії похибки спостережень. Перевірка гіпотез значущості параметрів множинної регресії. <i>Рекомендована література:</i> [5], с. 612–634.
1.2.1	<b>ОНК лінійної функції від вектора невідомих параметрів.</b> Критична область критерію Фішера, у якому використовується ОНК лінійної функції від параметра функції регресії. Технічна лема про $\chi^2$ -квадрат розподіл умовного мінімуму за лінійною гіпотезою функціонала найменших квадратів. <i>Рекомендована література:</i> [2], с. 6–14; [5], с. 624-648.
1.2.2	<b>Критерій Фішера перевірки лінійних гіпотез.</b> Перевірка гіпотези про відсутність корисних предикторів у множинній лінійній моделі регресії. Зведення до лінійної гіпотези та критерію Фішера. Таблиця ANOVA. <i>Рекомендована література:</i> [1], с. 39–60; [5], с. 612–621.
2.1	<b>Однофакторний ANOVA 1.</b> Постановка задачі однофакторного ANOVA. Вкладення цієї задачі у задачу перевірки лінійної гіпотези у множинній лінійній регресії. Порівняння двох середніх нормальних сукупностей за допомогою критерію Стьюдента та однофакторного ANOVA. <i>Рекомендована література:</i> [1], с. 13–28; 31-59; [2], с. 2–14; [3], с. 529-532; [5], с. 385-394.

2.2.1	<b>Однофакторний ANOVA 2.</b> Перевірка гіпотези рівності середніх декількох нормальних сукупностей методом однофакторного ANOVA. Перевірка гіпотези про рівність дисперсій декількох нормальних сукупностей. <i>Рекомендована література:</i> [2], с. 2–14; [3], с. 529-532; [4], с. 140-147; [5], с. 394-410.
2.2.2	<b>Однофакторний ANOVA 3.</b> Проблема Беренса-Фішера. Додатковий аналіз (post hoc analysis) вибірок після відхилення нульової гіпотези про рівність середніх у однофакторному ANOVA. <i>Рекомендована література:</i> [5], с. 397-405; [8], с. 1–16.
3.1.1	<b>Двофакторний ANOVA 1.</b> Класифікація даних за двома ознаками у випадку одного спостереження у кожній комірці. Критерій Тьюкі для виявлення неадитивності. <i>Рекомендована література:</i> [2], с. 15-23; [6], с. 122-130; [7], с. 218–224.
3.1.2	<b>Двофакторний ANOVA 2.</b> Класифікація даних за двома ознаками у випадку декількох, але однакового числа, спостережень у кожній комірці. Таж сама задача, але кількості спостережень в кожній комірці не однакові. <i>Рекомендована література:</i> [2], с. 15-23; [6], с. 130-140; [7], с. 224–228.
3.2.1	<b>Двофакторний ANOVA із взаємодією факторів.</b> Фіксовані та випадкові ефекти. Змішані ефекти. <i>Рекомендована література:</i> [2], с. 24-39; [5], с. 424-432; [6], с. 156–165.
3.2.2	<b>Загальна модель двофакторного ANOVA.</b> Модель компонент дисперсії та аналіз загальної моделі у припущенні, що всі спостереження є сумісно гауссівськими. <i>Рекомендована література:</i> [2], с. 24-39; [6], с. 141-145; [7], с. 228-234.
4.1	<b>Трифакторний ANOVA 1.</b> Головні ефекти та взаємодії у трифакторному ANOVA. Оцінювання головних ефектів та взаємодій для моделей з однаковою кількістю спостережень у кожній комірці. <i>Рекомендована література:</i> [7], с.145-150.
4.2	<b>Трифакторний ANOVA 2.</b> Ортогональні співвідношення у випадку рівних чисел спостережень у комірцях. Таблиця трифакторного ANOVA та пов'язані з нею обчислення. <i>Рекомендована література:</i> [3], с. 543-545; [4], с. 358-368; [7], с. 145-150.
5.1	<b>Рандомізований блок-дизайн.</b> Перевірка гіпотез. Надійний інтервал для різниці між середніми двох популяцій. <i>Рекомендована література:</i> [1], с. 315-324; [2], с. 41-62; [5], с. 413-424.
5.2.1	<b>ANOVA латинського квадрата як неповний трифакторний ANOVA.</b> Статистичний опис схеми латинського квадрата. Деякі обчислення. <i>Рекомендована література:</i> [1], с. 324-337; [2], с. 47-57; [4], с. 478-536; [6], с.175-187; [9], с. 62-70.
5.2.2	<b>Таблиця ANOVA латинського квадрата.</b> Питання теорії, пов'язані з побудовою таблиці. <i>Рекомендована література:</i> [1], с. 324-337; [2], с. 47-57; [4], с. 478-536; [6], с.175-187; [9], с. 62-70.
6.1.1	<b>Дробові репліки ПФЕ <math>2^m</math> (1).</b> Напіврепліки $2^{m-1}$ . Репліки $2^{m-p}$ . Генеруючі співвідношення. <i>Рекомендована література:</i> [9], с. 53-62.
6.1.2	<b>Дробові репліки факторного плану <math>2^m</math> (2).</b> Обчислення ОНК в ПФЕ $2^m$ та його репліках. <i>Рекомендована література:</i> [9], с. 53-62.

## Практичні заняття

№ з/п	Назва теми заняття та перелік основних питань (перелік дидактичних засобів, посилання на літературу та завдання на СРС)
1.	Коло задач, пов'язаних із множинною лінійною моделлю регресії. Критерій Фішера перевірки лінійних гіпотез. Таблиця ANOVA. Завдання СРС: [5], с. 605-635.
2.	Перевірка гіпотез рівності середніх двох та декількох нормальних вибірок в однофакторному ANOVA. Перевірка гіпотези про рівність дисперсій декількох нормальних сукупностей. Завдання СРС: [2], с. 2-14; [5], с. 385-394.
3.	Post hoc аналіз вибірок після відхилення нульової гіпотези про рівність середніх у однофакторному ANOVA. Завдання СРС: [5], с. 397-405.
4.	Класифікація даних у двофакторному ANOVA. Критерій Тьюкі для виявлення неадитивності. Завдання СРС: [2], с. 15-39; [6], с. 122-140.
5.	Двофакторний ANOVA із взаємодією факторів. Різні типи ефектів. Завдання СРС: [2], с.40-46; [5], с.424-432.
6.	Головні ефекти та взаємодії у трифакторному ANOVA. Таблиця ANOVA. Завдання СРС: [3], с.543-545; [4], с. 358-363; [7], с. 145-150.
7.	ANOVA латинського квадрата та відповідна таблиця. Завдання СРС: [2], с.47-57; [9], с.62-70.
8.	ПФЕ 2 <sup>m</sup> та його репліки. Обчислення ОНК. Завдання СРС: [9], с.53-62.
9.	Модульна контрольна робота.

### 6. Самостійна робота студента/аспіранта другого рівня вищої освіти

Завдання дисципліни включає наступні види самостійної роботи:

- підготовка до лекційних та практичних занять;
- виконання модульної контрольної роботи;
- виконання розрахунково-графічної роботи;
- підготовка до іспиту.

### ● Політика та контроль

### 7. Політика навчальної дисципліни (освітнього компонента)

**Рекомендовані методи навчання:** вивчення основної та допоміжної літератури за тематикою лекцій, розв'язування задач на практичних заняттях. Здобувачу другого рівня вищої освіти рекомендується вести докладний конспект лекцій. Важливою частиною якісного засвоєння навчального матеріалу та відпрацювання методів розв'язання основних завдань дисципліни є самостійна робота. Крім казаного вище, вона містить підготовку до МКР, РГР та екзамену.

## Академічна доброчесність

Політика та принципи академічної доброчесності визначені у розділі 3 Кодексу честі Національного технічного університету України «Київський політехнічний інститут імені Ігоря Сікорського». Детальніше: <https://kpi.ua/code>

## Норми етичної поведінки

Норми етичної поведінки студентів і працівників визначені у розділі 2 Кодексу честі Національного технічного університету України «Київський політехнічний інститут імені Ігоря Сікорського». Детальніше: <https://kpi.ua/code>

### 8. Види контролю та рейтингова система оцінювання результатів навчання (PCO)

На першому занятті здобувачі ознайомлюються із рейтинговою системою оцінювання (PCO) дисципліни, яка побудована на основі Положення про систему оцінювання результатів навчання [https://document.kpi.ua/files/2020\\_1-273.pdf](https://document.kpi.ua/files/2020_1-273.pdf).

Зокрема, рейтинг здобувача третього рівня вищої освіти з освітнього компонента формується як сума балів поточної успішності навчання – стартового рейтингу (максимально 50 балів) та балів за іспит (максимально 50 балів).

Поточний контроль: фронтальний (усний, письмовий), МКР, РГР.

Календарний контроль: проводиться двічі на семестр як моніторинг поточного стану виконання вимог силабусу.

Рейтингова система оцінювання включає якість виконання домашніх завдань МКР та РГР. Кожний здобувач другого рівня освіти отримує свій підсумковий рейтинг з дисципліни.

Рейтинг здобувача другого рівня освіти з кредитного модуля складається з балів, які він отримує за:

- виконання домашніх завдань;
- написання модульної контрольної роботи;
- виконання розрахунково-графічної роботи;
- відповіді на екзамені.

### Система рейтингових (вагових) балів та критеріїв оцінювання:

Метод оцінювання	Кількість	Мінімальна оцінка в балах	Максимальна оцінка в балах
Виконання домашніх завдань	5	5	10
Модульна контрольна робота	1	12	20
Розрахунково-графічна робота	1	13	20
Стартовий рейтинг		30	50
Іспит	1		50
Підсумковий		60	100

**Сума стартових балів та балів за іспит переводиться до екзаменаційної оцінки згідно з таблицею:**

100...95	Відмінно
94...85	Дуже добре
84...75	Добре
74...65	Задовільно
64...60	Достатньо
Менше 60	Незадовільно
Стартовий рейтинг менше 30 балів	Не допущено

**Робочу програму навчальної дисципліни (силабус):**

**Складено** професором кафедри МА та ТЙ, д.ф.-м.н., професором Івановим О.В.

**Ухвалено** кафедрою математичного аналізу та теорії ймовірностей (протокол № 12 від 19.06.2023 р.)

**Погоджено** Методичною комісією ФМФ (протокол № 10 від 27.06.2023 р.)