

Міністерство освіти і науки України
Національний університет водного господарства та природокористування
Навчально-науковий інститут економіки та менеджменту

Кафедра економічної кібернетики

ЗАТВЕРДЖУЮ

Проректор з науково-педагогічної,
методичної та виховної роботи

_____ О.А. Лагоднюк

“ ____ ” _____ 2019 р.

06-11-43

РОБОЧА ПРОГРАМА НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

Program of the Discipline

АНАЛІЗ ДАНИХ

DATA ANALYSIS

Спеціальність:

Specialty

051 Економіка

051 Economics

Освітні програми: Економічна кібернетика, Бізнес аналітика, Управління персоналом і економіка праці

Specialization Economic Cybernetics, Business Analytics, Human Resources Management and Labor Economics

Робоча програма дисципліни «Аналіз даних» для здобувачів вищої освіти I (бакалаврського) рівня за спеціальністю 051 «Економіка», ОПП «Економічна кібернетика», «Бізнес аналітика», «Управління персоналом і економіка праці». – Рівне: НУВГП, 2019. – 14 с.

Розробник: П.М. Грицюк, завідувач кафедри економічної кібернетики, д.е.н., професор

Робочу програму схвалено на засіданні кафедри економічної кібернетики

Протокол від " 27 " грудня 2018 року № 08

Завідувач кафедри
економічної кібернетики _____ (П.М. Грицюк)
(підпис)

Схвалено науково-методичною комісією за спеціальністю 051 «Економіка»

Протокол від " 31 " січня 2019 року № 01

Голова науково-методичної комісії _____ Безтелесна Л.І.
(підпис)

ВСТУП

Програма навчальної дисципліни «Аналіз даних» складена на основі ОПП та навчального плану підготовки бакалаврів за спеціальністю 051 «Економіка» ОПП «Економічна кібернетика», «Бізнес аналітика», «Управління персоналом і економіка праці».

Предметом вивчення навчальної дисципліни є формування теоретичних знань щодо встановлення однорідності та закону розподілу числових вибірок, методики проведення кореляційного, регресійного, дисперсійного та факторного аналізу даних; засвоєння практичних навичок графічного аналізу даних, перевірки статистичних гіпотез, кореляційного, регресійного, дисперсійного та факторного аналізу даних.

Міждисциплінарні зв'язки: дисципліна «Аналіз даних» є складовою частиною циклу загальної підготовки для здобувачів вищої освіти першого (бакалаврського) рівня за спеціальністю 051 «Економіка». Вивчення дисципліни передбачає наявність систематичних та ґрунтовних знань із суміжних курсів – «Вища математика», «Економічна інформатика», «Математична статистика», «Мікроекономіка» та дисциплін, які безпосередньо формують компетенції фахівця відповідної спеціальності, цілеспрямованої роботи над вивченням спеціальної літератури, активної роботи на лекціях, лабораторних заняттях, самостійної роботи над виконанням поставлених завдань.

Вимоги до знань та умінь визначаються галузевими стандартами вищої освіти України.

Анотація

Відмінною рисою сучасної економіки є її математизація та комп'ютеризація. Прийняття обґрунтованих рішень в економіці неможливе без попереднього системного та статистичного аналізу економічних явищ і процесів. Економіко-статистичні методи широко застосовуються для вирішення таких задач, як розрахунок інвестиційних бізнес-проектів, маркетингові дослідження, оптимізація промислового та аграрного виробництва, логістичних процесів, оцінювання економічних ризиків тощо. Для комп'ютерного вирішення статистичних задач створені сучасні програми та програмні комплекси. Найбільш відомими з них є: пакет STATISTICA, пакет SPSS (Statistical Package for Social Science), пакет STATA, мова програмування R.

Вивчення даної дисципліни має за мету озброїти здобувачів вищої освіти сучасними методами обробки, аналізу та класифікації даних. Вивчення дисципліни передбачає розгляд таких розділів, як: випадкові величини та закони розподілу, методи перевірки гіпотез, кореляційний, дисперсійний та регресійний аналіз, факторний аналіз, методи багатовимірного аналізу статистичних даних.

Предмет дисципліни: теорія математичної статистики та обробки даних.

Ключові слова: вибірка, випадкова величина, гіпотеза, дисперсія, закон розподілу, класифікація, кореляція, регресія.

Abstract

A distinctive feature of the modern economy is its mathematization and computerization. Making the reasonable decisions in the economy is impossible without a preliminary system and statistical analysis of economic phenomena and processes. Economic and statistical methods are widely used for solving such tasks as calculation of investment business projects, marketing researches, optimization of industrial and agricultural production, logistic processes, estimation of economic risks, etc. Modern software and software complexes are created for the computer solution of statistical tasks. The most popular of these are: STATISTICA package, SPSS (Statistical Package for Social Science) package, STATA package, programming language R.

The study of this discipline aims at equipping higher education graduates with modern methods of processing, analysis and classification of data. Studying discipline involves consideration of such sections as: random variables and distribution laws, hypothesis testing methods, correlation, dispersion and regression analysis, factor analysis, methods of multidimensional analysis of statistical data.

The subject: the theory of mathematical statistics and data processing.

Keywords: sampling, random variable, hypothesis, variance, distribution law, classification, correlation, regression.

1. Опис навчальної дисципліни

Найменування показників	Галузь знань, спеціальність, спеціалізація, рівень вищої освіти	Характеристика навчальної дисципліни	
		денна форма навчання	заочна форма навчання
Кількість кредитів: 4	Галузь знань: 05 Соціальні та поведінкові науки	Цикл загальної підготовки	
Модулів: 1	Спеціальність: 051 Економіка	<i>Рік підготовки</i>	
Змістових модулів: 2	Спеціалізація: Економічна кібернетика, Бізнес аналітика, Управління персоналом і економіка праці	2-й	1-й
	Рівень вищої освіти: I бакалаврський	<i>Семестр</i>	
Загальна кількість годин – 120		4-й	1-й
Тижневих годин для денної форми навчання:		<i>Лекції</i>	
		22 год.	6 год.
		<i>Лабораторні роботи</i>	
		20 год.	6 год.
аудиторних – 4		<i>Самостійна робота</i>	
		78 год.	108 год.
самостійної роботи студента – 6		<i>Індивідуальні завдання:</i>	
		-	-
		<i>Вид контролю:</i>	
		Залік	Залік

Примітка.

Співвідношення кількості годин аудиторних занять і самостійної роботи становить:

для денної форми навчання – 35% до 65%;

для заочної форми навчання – 8% до 92%.

2. Мета та завдання навчальної дисципліни

Метою вивчення навчальної дисципліни "Аналіз даних" є формування навиків виконання послідовних дій з інтерпретації зібраних даних та їх перетворення у статистичні форми, необхідні для ухвалення маркетингових та управлінських рішень.

Завдання вивчення дисципліни полягає у засвоєнні теоретичних основ та виробленні практичних навиків класифікації експериментальних даних, пошуку закономірностей та залежностей між змінними з використанням засобів комп'ютерних технологій.

У результаті вивчення навчальної дисципліни здобувач вищої освіти повинен **знати:**

- методи класифікації даних;
- методи аналізу вибіркового даних;
- методи пошуку закономірностей у даних та залежностей між випадковими величинами;
- методи перевірки статистичних гіпотез.

Здобувач вищої освіти повинен **вміти:**

- застосовувати сучасні пакети прикладних програм для перевірки гіпотез та аналізу даних;
- проводити попередній (розвідковий) аналіз даних та графічний аналіз даних;
- здійснювати класифікацію (кластеризацію) даних;
- здійснювати кореляційно-регресійний аналіз для пошуку залежностей між випадковими величинами;
- перевіряти гіпотезу про нормальний розподіл даних.

Програма розрахована на здобувачів вищої освіти, які навчаються за освітньо-кваліфікаційними програмами підготовки бакалаврів.

3. Програма навчальної дисципліни

ЗМІСТОВИЙ МОДУЛЬ 1. Попередня обробка даних. Перевірка гіпотез

Тема 1. Основні поняття та основні задачі аналізу даних

Мета і завдання дисципліни. Етапи аналізу даних. Класифікація типів змінних. Основні методи аналізу даних. Попередній аналіз даних. Генеральна сукупність і вибірка.

Тема 2. Зведення та групування статистичних даних

Методи групування статистичних даних. Варіаційний ряд. Емпірична функція розподілу. Встановлення групових інтервалів. Методи побудови статистичних таблиць.

Тема 3. Попередня обробка даних

Етапи попередньої обробки даних. Числові характеристики вибірки. Дискретний статистичний розподіл. Інтервальний статистичний розподіл. Емпіричні моменти. Видалення аномальних спостережень.

Тема 4. Нормальний закон розподілу

Нормальний розподіл даних. Щільність розподілу. Інтегральна функція розподілу. Правило 3σ . Перевірка гіпотези про нормальний розподіл. Критерій Пірсона. Критерій Колмогорова-Смірнова. Тест Жака-Бера. Оцінка ймовірності випадкових подій.

Тема 5. Статистична перевірка гіпотез

Основна і альтернативна гіпотеза. Рівень значущості. Задачі першого і другого типу. Перевірка гіпотези про рівність дисперсій. Перевірка гіпотези про рівність середнього заданому рівню A . Перевірка гіпотези про залежність середнього від експерименту. Порівняння середніх двох вибірок з однаковими дисперсіями. Порівняння середніх двох вибірок з різними дисперсіями.

ЗМІСТОВИЙ МОДУЛЬ 2. Статистичне вивчення зв'язків між факторами. Класифікація даних

Тема 6. Статистичне вивчення зв'язків між факторами

Функціональний і стохастичний зв'язок між випадковими величинами. Кореляція факторів. Критерій істотності кореляції. Парна лінійна регресія. Нелінійна регресія. Множинна регресія. Кореляційний аналіз множинної регресії.

Тема 7. Рангова кореляція

Рангові змінні. Коефіцієнт Спірмена. Коефіцієнт рангової кореляції Кендала. Експертне оцінювання зразків. Коефіцієнт конкордації. Бісеріальна кореляція.

Тема 8. Кореляційний аналіз номінальних даних

Кростабуляція даних. Коефіцієнт асоціації. Таблиця контингенції. Коефіцієнт спряженості.

Тема 9. Дисперсійний аналіз

Моделі дисперсійного аналізу. Однофакторний дисперсійний аналіз. Критерій Фішера. Критерій Бартлета. Двохфакторний дисперсійний аналіз.

Тема 10. Інші методи аналізу

Факторний аналіз. Метод головних компонент. Розмірність інформаційного потоку. Класифікація даних. Дискримінантний аналіз. Кластерний аналіз.

4. СТРУКТУРА НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

Назви змістових модулів і тем	Кількість годин							
	денна форма				заочна форма			
	усього	у тому числі			усього	у тому числі		
		л	лаб.	с. р.		л	лаб.	с. р.
1	2	3	4	5	6	7	8	9
Змістовий модуль 1. Попередня обробка даних. Перевірка гіпотез								
Тема 1. Основні поняття та основні задачі аналізу даних	8	2	0	6	12	-		8
Тема 2. Зведення та групування статистичних даних	12	2	2	8	12	1	1	12
Тема 3. Попередня обробка даних. Емпірична функція розподілу	12	2	2	8	18	1	1	12
Тема 4. Нормальний закон розподілу	12	2	2	8	20	1	1	12
Тема 5. Статистична перевірка гіпотез	12	2	2	8	18	-	-	12
Всього змістовий модуль 1	56	10	8	38	80	3	3	56
Змістовий модуль 2. Статистичне вивчення зв'язків між факторами. Класифікація даних								
Тема 6. Статистичне вивчення зв'язків між факторами	18	4	4	10	22	1	1	12

Тема 7. Рангова кореляція. Коефіцієнт Спірмена. Коефіцієнт Кендала	12	2	2	8	12	1	1	10
Тема 8. Кореляційний аналіз номінальних даних	12	2	2	8	12	1	1	10
Тема 9. Однофакторний та двох- факторний дисперсійний аналіз	12	2	2	8	12	-	-	10
Тема 10. Факторний аналіз. Дискримінантний аналіз. Кластерний аналіз	10	2	2	6	12	-	-	10
Всього змістовий модуль 2	64	12	12	40	70	3	3	52
Усього годин	120	22	20	78	150	6	6	108

5. ТЕМИ ЛАБОРАТОРНИХ РОБІТ

№ з/п	Назва теми	Кількість годин	
		Денна форма	Заочна форма
1	Графічний аналіз даних	2	
2	Попередня обробка даних	2	1
3	Перевірка гіпотези про нормальний розподіл	2	1
4	Перевірка статистичних гіпотез про середнє та дисперсію	2	1
5	Парна лінійна регресія	2	
6	Множинна лінійна регресія	2	1
7	Однофакторний дисперсійний аналіз	2	1
8	Кластерний аналіз	2	1
9	Дискримінантний аналіз	2	
10	Підсумкове заняття	2	
	Усього годин	20	6

6. САМОСТІЙНА РОБОТА

Розподіл годин самостійної роботи для студентів денної форми навчання:
 Підготовка до аудиторних занять – 0,5 год/1 год. Занять 42*0,5 = 21 год.
 Підготовка до контрольних заходів – 6 год на 1 кредит ЄКТС. 6*4 = 24 год.

Разом 45 год. Загальний обсяг самостійної роботи 78 год., відповідно на самостійне опрацювання виділяється 33 год.

6.1. Завдання для самостійної роботи

№ з/п	Тема самостійної роботи	Кількість годин
1	Тема 1. Пакет Statistica. Перевірка гіпотези про нормальний розподіл засобами пакета Statistica. Критерії Колмогорова-Смірнова та Шапіро-Вілка	6
2	Тема 2. Інструмент статистика у Microsoft Excel. Кореляційно-регресійний аналіз в середовищі Microsoft Excel	6
3	Тема 3. Метод головних факторів. Розмірність інформаційного потоку. Визначення розмірності засобами Matlab	5
4	Тема 4. Класифікація даних. Класифікація без навчання. Класифікація з навчанням.	5
5	Тема 5. Перевірка статистичних гіпотез. Непараметрична статистика	5
6	Тема 6. Ідентифікація закону розподілу вибірки. Логнормальний закон розподілу. Показниковий закон розподілу.	6
	Разом	42

Звіт про самостійну роботу подається у вигляді конспекту з кожної теми, наведеної у таблиці 6.1. Звіт оформляється в окремому зошиті для самостійної роботи, або у вигляді додатків до конспекту лекція.

7. Методи навчання

При викладанні навчальної дисципліни «Аналіз даних» використовується інформаційно-ілюстративний та проблемно-активізуючий методи навчання.

Лекції проводяться у супроводі мультимедійних презентацій з основними теоретичними положеннями теми та прикладами розв'язування задач.

Лабораторні заняття проводяться в комп'ютерному класі. При цьому виконується індивідуальне завдання, оформляється і захищається звіт про виконану роботу. Для виконання завдань використовуються сучасне програмне математичне забезпечення: MS Excel, Statistica, EViews.

8. Методи контролю

Поточний контроль знань студентів з навчальної дисципліни проводиться у письмовій формі та шляхом усного опитування. Проміжний контроль знань студентів проводиться двічі на семестр у вигляді комп'ютерного тестування. Контрольні завдання за змістовим модулем включають тестові питання (200 тестів, одна правильна відповідь з п'яти запропонованих).

Контроль самостійної роботи проводиться:

- з лекційного матеріалу – шляхом перевірки конспектів;
- з лабораторних занять – шляхом перевірки виконаних завдань та оформлених звітів.

Усі форми контролю включено до 100-бальної шкали оцінки.

Оцінювання результатів поточної роботи (завдань, що виконуються на лабораторних заняттях, результати самостійної роботи студентів) проводиться за такими критеріями:

1. Розрахункові завдання, задачі, лабораторні роботи (у % від кількості балів, виділених на завдання із заокругленням до цілого числа):

0 % – завдання не виконано;

40% – завдання виконано частково та містить суттєві помилки методичного або розрахункового характеру;

60% – завдання виконано повністю, але містить суттєві помилки у розрахунках або в методиці;

80% – завдання виконано повністю і вчасно, проте містить окремі несуттєві недоліки (розмірності, висновки, оформлення тощо);

100% – завдання виконано правильно, вчасно і без зауважень.

9. Розподіл балів, які отримують студенти

Шкала оцінювання студентів денної форми навчання

Поточне тестування та самостійна робота										Сума
Змістовий модуль 1					Змістовий модуль 2					
T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	
9	11	11	11	11	11	9	9	9	9	100

T1, T2...T10 – теми змістових модулів.

У заліково-екзаменаційній відомості результати навчання проставляються за двома шкалами – 100-бальною та національною. Позитивні оцінки виставляються тільки тим студентам, які виконали всі види навчальної роботи, передбачені робочою програмою навчальної дисципліни, і набрали за результатами поточного контролю не менше 60 балів.

Шкала оцінювання

Сума балів за всі види навчальної діяльності	Оцінка за національною шкалою
	Для заліку
90-100	зараховано
82-89	
74-81	
64-73	
60-63	
35-59	не зараховано з можливістю повторного складання
0-34	не зараховано з обов'язковим повторним вивченням ди-

10. Методичне забезпечення

Методичне забезпечення навчальної дисципліни „Аналіз даних”

- інтерактивний комплекс навчально-методичного забезпечення дисципліни;
- конспект лекцій на паперовому носії;
- конспект лекцій на електронному носії;
- комплект мультимедійних презентацій;
- методичні рекомендації до виконання лабораторних робіт з дисципліни «Аналіз даних», розміщені на електронних носіях;
- пакети тестових завдань по кожній темі і в цілому по всьому курсу.

11. РЕКОМЕНДОВАНА ЛІТЕРАТУРА

Базова

1. Айвазян С. А., Енюков И. С., Мешалкин Л. Д. Прикладная статистика: основы моделирования и первичная обработка данных. Москва: Финансы и статистика, 1983.
2. Бахрушин В.Є. Методи аналізу даних : навч. посібник / В.Є. Бахрушин, Запоріжжя : КПУ, 2011, 268 с.
3. Лапач С.Н., Чубенко А.В., Бабич П.Н. Статистика в науке и бизнесе. Киев: МОРИОН, 2002.
4. Макарова Н.В., Трофимец В.Я. Статистика в Excel. Москва: Финансы и статистика, 2002.
5. Сільченко М.В. Інформатика. Комп'ютерний аналіз економічних даних: моніторинг знань: зб. практ. завдань / М. В. Сільченко, Т. О. Кучерява, Ю. М. Красюк ; за заг. ред. О. Д. Шаропова. Київ: КНЕУ, 2013, 354 с.

6. Ферстер Э., Ренц Б. Методы корреляционного и регрессионного анализа. Москва: Финансы и статистика, 1983.

Допоміжна

1. Андерсон Т. Статистический анализ временных рядов. Москва, Мир, 1976.
2. Аронов В.И. Методы математической обработки данных на ЭВМ. Москва, Недра, 1977.
3. Кендалл М. Дж., Стюарт А. Многомерный анализ. Москва, Наука, 1976.
4. Дж. Вен Райзман. Классификация и кластеры. Москва, Мир, 1980.
5. Шеффе Г. Дисперсионный анализ. Изд. 2-е. Москва, Наука, 1980.
6. Гутер Р.С., Овчинский Б. В. Элементы численного анализа и математической обработки данных эксперимента. Москва, 1971.

Електронний репозиторій НУВГП

1. Грицюк П.М. (2017) Аналіз даних [Опис дисципліни] [Електронний ресурс]. – Режим доступу: <http://ep3.nuwm.edu.ua/id/eprint/13333>

12. ІНФОРМАЦІЙНІ РЕСУРСИ

1. Наукова бібліотека НУВГП (м. Рівне, вул. Олекси Новака, 75). – Режим доступу: <http://lib.nuwm.edu.ua>, <http://ep3.nuwm.edu.ua>.
2. Національна бібліотека ім. В.І. Вернадського. – Режим доступу: <http://www.nbuv.gov.ua/>
3. Рівненська обласна універсальна наукова бібліотека. - м. Рівне, майдан Короленка, 6. – Режим доступу : <http://www.lib.rv.ua/>
4. Державний комітет статистики України. – режим доступу: <http://www.ukrstat.gov.ua/>
5. <http://datan.ucoz.ru> (сайт циклу дисциплін “Аналіз даних”, “Математичні методи аналізу даних”, “Комп’ютеризовані технології аналізу даних”)
6. <http://www.basegroup.ru> (сайт BaseGroup Labs – провідного російського розробника програмного забезпечення з аналізу даних)
7. <http://www.statsoft.ru/home/textbook/default.htm> (електронний підручник статистики фірми StatSoft)
8. <http://orlovs.pp.ru> (сайт “Високі статистичні технології” професора О.І. Орлова)
9. <http://www.aup.ru/books/m163/> (підручник О.І. Орлова “Прикладна статистика”)
10. <http://www.ami.nstu.ru/~headrd> (сторінка професора Б.Ю. Лемешка на сайті Новосибірського державного технічного університету)

