

НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ БІОРЕСУРСІВ І ПРИРОДОКОРИСТУВАННЯ УКРАЇНИ

Кафедра економічної кібернетики



«ЗАТВЕРДУЮ»

декан факультету інформаційних технологій

Глазунова О.Г.

2023 р.

«СХВАЛЕНО»

Протокол №10 від 18.05.2023 р.

Завідувач кафедри

Жерліцин Д.М.

«Розглянуто»

Гарант ОП «Інформаційні системи і технології»

Смолій В.М.

РОБОЧА ПРОГРАМА НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

«Аналіз і візуалізація даних»

Галузь знань 12 «Інформаційні технології»
Спеціальність 126 «Інформаційні системи і технології»
Освітня програма «Інформаційні системи і технології»
Факультет інформаційних технологій

Розробник:

доцент кафедри економічної кібернетики, д.е.н., доц. Кравченко В.М.

Київ – 2023 р.

1. Опис навчальної дисципліни

Аналіз і візуалізація даних

Галузь знань, напрям підготовки, спеціальність, ступень вищої освіти	
Ступень вищої освіти	<i>бакалавр</i>
Галузь знань	12 Інформаційні технології
Спеціальність	126 Інформаційні системи і технології
Освітня програма	Інформаційні системи і технології
Характеристика навчальної дисципліни	
Вид	Обов'язкова
Загальна кількість годин	150
Кількість кредитів ECTS	5
Кількість змістовних модулів	2
Курсовий проект (робота) (за наявності)	Не передбачено
Форма контролю	Іспит
Показники навчальної дисципліни	
Форма навчання	<i>Денна</i>
Рік підготовки	<i>2</i>
Семестр	<i>4</i>
Лекційні заняття	<i>30</i>
Практичні, семінарські заняття	
Лабораторні заняття	<i>30</i>
Самостійна робота студента	<i>90</i>
Індивідуальні завдання	
Кількість тижневих аудиторних годин для денної форми навчання	<i>4 години на тиждень</i>

2. Мета, завдання та компетентності навчальної дисципліни

Протягом курсу розглядаються основні принципи збирання й оброблення даних, поняття відкритих даних, особливості аналітики в економіці, бізнесі, сферах природокористування. Значна увага приділяється ключовим підходам до візуалізації даних, етапи попередньої і тематичної обробки даних. Практична частина курсу передбачає опанування аналітичних інструментів MS Excel, а також мов програмування Python і R в контексті обробки первинної обробки даних, реалізації методів описової статистики, групування та очищення даних, аналізу часових рядів, інтерактивних методів аналізу і обґрунтування рішень на практиці. Передбачається створення аналітичних звітів (презентацій, інфографіки), формування та оприлюднення портфолію аналітика.

У результаті вивчення дисципліни студент набуває здібностей: збирати та проводити первинне оброблення великих обсягів даних з різних джерел походження; проводити аналітичну обробку інформації різної структури та змісту; використовувати табличний редактор, а також можливості мов програмування Python і R для аналізу та візуалізації даних; автоматизувати процес підготовки аналітичних звітів та інфографічних матеріалів.

Мета вивчення курсу – оволодіння фундаментальними знаннями щодо аналізу емпіричних даних, їх візуалізації та підготовки аналітичних звітів з використанням табличного редактора MS Excel, мов програмування Python і R.

Завдання вивчення курсу

Засвоївши курс студент повинен:

знати:

теоретичні основи та принципи аналізу даних;
етапи проведення статистичного аналізу даних;
особливості порівняння груп даних;
основні статистичні моделі та методи вимірювання взаємозв'язків;
особливості та структуру часових рядів даних;
принципи побудову статистичних таблиць та графіків;
особливості застосування інструментів аналізу та візуалізації даних у MS Excel і додатках з Python і R;

вміти:

розраховувати та інтерпретувати показники описової статистики;
перевіряти статистичні гіпотези;
розраховувати показники інтенсивності динаміки даних часових рядів;

будувати та аналізувати різні види графіків (нестандартні, комбіновані);
використовувати данні з різних джерел для побудови діаграм;
виконувати сценарний аналіз і обґрунтовувати управлінські рішення;

володіти:

вибірковим методом статистичного аналізу даних,
методом дисперсійного аналізу;
статистичними моделями вимірювання взаємозв'язків;
індексним методом аналізу часових рядів даних;
методами прогнозування часових рядів даних;
інструментами графічного аналізу статистичної інформації;
інструментами аналізу та візуалізації даних з MS Excel;
інструментами сценарного аналізу та підтримки прийняття рішень з MS Excel.

Набуття компетентностей:

загальні компетентності (ЗК):

- ЗК1. Здатність до абстрактного мислення, аналізу та синтезу.
- ЗК2. Здатність застосовувати знання у практичних ситуаціях.
- ЗК3. Знання та розуміння предметної області та розуміння професійної діяльності.
- ЗК6. Здатність вчитися й оволодівати сучасними знаннями.
- ЗК7. Здатність до пошуку, оброблення та аналізу інформації з різних джерел.
- ЗК8. Здатність генерувати нові ідеї (креативність).
- ЗК11. Здатність приймати обґрунтовані рішення.
- ЗК12. Здатність оцінювати та забезпечувати якість виконуваних робіт.

фахові (спеціальні) компетентності (СК):

- СК1. Здатність аналізувати об'єкт проектування або функціонування та його предметну область.
- СК2. Здатність застосовувати стандарти в області інформаційних систем та технологій при розробці функціональних профілів, побудові та інтеграції систем, продуктів, сервісів і елементів інфраструктури організації.
- СК7. Здатність застосовувати інформаційні технології у ході створення, впровадження та експлуатації системи менеджменту якості та оцінювати витрати на її розроблення та забезпечення.

Програмні результати:

ПР2. Застосовувати знання фундаментальних і природничих наук, системного аналізу та технологій моделювання, стандартних алгоритмів та дискретного аналізу при розв'язанні задач проектування і використання інформаційних систем та технологій.

ПР4. Проводити системний аналіз об'єктів проектування та обґрунтовувати вибір структури, алгоритмів та способів передачі інформації в інформаційних системах та технологіях.

3. Програма та структура навчальної дисципліни

Змістовний Модуль 1. Статистичний аналіз даних і візуалізація його результатів

Тема 1. Основи статистичного аналізу й візуалізації даних.

Теоретичні основи статистичного аналізу даних. Особливості статистичного спостереження та вибіркового методу. Ключові джерела доступу до статистичних даних. Попередня обробка даних (пропуски, викиди). Сучасне ПЗ статистичного аналізу даних. Зведення та групування статистичних даних. Базові типи діаграм для візуалізації даних.

Тема 2. Статистичні показники.

Сутність і види статистичних показників, їх властивості. Абсолютні величини. Відносні величини. Середні величини й їх види (арифметична, гармонічна, хронологічна, геометрична тощо). Система статистичних показників.

Тема 3. Ряди розподілу. Аналіз варіацій та форми розподілу.

Закономірність розподілу. Характеристики центра розподілу. Характеристики варіації. Характеристики форми розподілу. Види та взаємозв'язок дисперсій. Функції MS Excel з обробки й аналізу даних: vlookup, hlookup, frequency, average, median, mode, var, stdev, min, max, range, avedev, countblank, count, quartile, percentile. «Описова Статистика» у пакеті "Аналіз Даних" MS Excel. Міжквартильний інтервал. Ящик з вусами (Box and Whisker Plot).

Тема 4. Вибірковий метод. Статистична перевірка гіпотез.

Суть вибіркового спостереження. Вибіркові оцінки середньої та частки. Різновиди вибірок. Визначення обсягу вибірки. Одно- і двовибіркові t-критерій та z-критерій. t-критерій для парних вибірок. z-критерій однієї та двох часток. Реалізація критеріїв перевірки гіпотез в Excel.

Змістовний модуль 2. Методи аналізу даних в процесі прийняття рішень

Тема 5. Дисперсійний аналіз (ANOVA).

Однофакторний дисперсійний аналіз. Довірчий інтервал. Ранговий однофакторний аналіз Краскела-Уолліса. Критерій Джонкхієра (Джонкхієра – Терпстра). М-критерій Бартлетта. G-критерій Кокрена (Кочрена). Критерій Левене. Критерій Брауна-Форсайта. Двофакторний дисперсійний аналіз. Ранговий критерій Фрідмана. Критерій Пейджа. Q-критерій Кокрена.

Тема 6. Аналіз взаємозв'язків. Кореляційно-регресійний аналіз.

Види взаємозв'язків. Оцінка щільності та перевірка істотності кореляційного зв'язку. Кореляція Пірсона Рангова кореляція Спірмена. Оцінка узгодженості варіації атрибутивних ознак. Регресійний аналіз: проста і множинна лінійна регресія. Оцінка регресійної моделі: метрики MSE, MAE, MAPE, R2 тощо.

Тема 7. Ряди динаміки. Індекси.

Суть і складові елементи динамічного ряду. Характеристики інтенсивності динаміки. Середня абсолютна та відносна швидкість розвитку. Характеристика основної тенденції розвитку. Оцінка коливань та сталості динаміки. Суть і функції індексів. Методологічні основи побудови зведених індексів. Агрегатна форма індексів. Середньозважені індекси. Взаємозв'язки індексів. Індекси середніх величин. Територіальні індекси.

Тема 8. Сценарний аналіз і пошук рішення в Excel.

Сценарний аналіз ("Що-якщо"): одно- та багатопродуктова модель прибутку. Аналіз еластичності, аналіз компромісів. Рівень беззбитковості, рівень байдужості. «Таблиця даних» (підстановок) Excel. «Підбір параметру» Excel. «Пошук рішення» / «Розв'язувач» (лінійна оптимізація) Excel.

Назви змістових модулів і тем	Кількість годин					
	усього	у тому числі				
		л	п	лаб.	інд.	с. р.
1	2	3	4	5	6	7
Змістовий модуль 1. Статистичний аналіз даних і візуалізація його результатів						
Тема 1. Основи статистичного аналізу й візуалізації даних	18	4	0	4	0	10
Тема 2. Статистичні показники	14	2	0	2	0	10
Тема 3. Ряди розподілу. Аналіз варіацій та форми розподілу	20	4	0	4	0	12
Тема 4. Вибірковий метод. Статистична перевірка гіпотез	20	4	0	4	0	12
Разом за змістовим модулем 1	72	14	0	14	0	44
Змістовний модуль 2. Методи аналізу даних в процесі прийняття рішень						
Тема 5. Дисперсійний аналіз (ANOVA)	20	4	0	4	0	12
Тема 6. Аналіз взаємозв'язків. Кореляційно-регресійний аналіз	20	4	0	4	0	12
Тема 7. Ряди динаміки. Індекси	18	4	0	4	0	10
Тема 8. Сценарний аналіз і пошук рішення в Excel	20	4	0	4	0	12
Разом за змістовим модулем 2	78	16	0	16	0	46
Усього годин	150	30	0	30	0	90

6. Теми лабораторних занять

№ з/п	Назва теми	Кількість годин
1.	Збирання та первинний аналіз даних	2
2.	Показники описової статистики та базові графічні методи аналізу даних	4
3.	Ряди розподілу та аналіз варіацій даних	4
4.	Статистична перевірка гіпотез	4
5.	Дисперсійний аналіз даних (ANOVA)	4
6.	Аналіз взаємозв'язків. Кореляційно-регресійний аналіз	4
7.	Індексний метод та прогнозування часових рядів даних	4
8.	Сценарний аналіз і пошук рішення в Excel	4
	Разом	30

7. Теми самостійної роботи

№ з/п	Назва теми	Кількість годин
1	Збирання та очищення даних з різних джерел з Python і R	45
2	Аналіз та візуалізація часових рядів й інших наборів даних з Python і R	45

8. Контрольні питання, комплекти тестів для визначення рівня засвоєння знань студентами.

Перелік питання для визначення рівня засвоєння знань студентами

1. Абсолютні величини під час аналізу статистичних показників
2. Агрегатна форма індексів
3. Взаємозв'язки індексів
4. Вибіркові оцінки середньої та частки
5. Види взаємозв'язків статистичних даних
6. Види та взаємозв'язок дисперсій статистичних даних
7. Визначення обсягу вибірки
8. Відносні величини під час аналізу статистичних показників
9. Закономірність розподілу статистичних даних
10. Збирання та попередній аналіз статистичної інформації
11. Індекси середніх величин
12. Класифікації та групування статистичних даних
13. Ключові показники описової статистики та елементарна графіка
14. Методологічні основи побудови зведених індексів
15. Особливості статистичного спостереження та вибіркового методу
16. Оцінка коливань та сталості динаміки у межах динамічного ряду
17. Оцінка узгодженості варіації атрибутивних ознак
18. Принципи побудови та аналізу зведених даних в Excel.
19. Принципи побудови та аналізу зведених даних за допомогою сучасних інструментів ІТ та програмних засобів
20. Рангова кореляція
21. Різновиди статистичних вибірок
22. Середні величини під час аналізу статистичних показників
23. Середньозважені індекси
24. Середня абсолютна та відносна швидкість розвитку в межах динамічного ряду
25. Система статистичних показників під час аналізу статистичних показників
26. Статистична перевірка гіпотез
27. Сутність і види статистичних показників

28. Сутність статистичного зведення та принципи побудови статистичних таблиць
29. Суть вибіркового спостереження
30. Суть і складові елементи динамічного ряду
31. Суть і функції статистичних індексів
32. Сучасне програмне забезпечення статистичного аналізу даних
33. Теоретичні основи статистичного аналізу даних та поняття сучасної статистики
34. Характеристики варіації статистичних даних.
35. Характеристики інтенсивності динаміки у межах динамічного ряду
36. Характеристики центра розподілу статистичних даних

Приклад екзаменаційного білету

НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ БІОРЕСУРСІВ І ПРИРОДОКОРИСТУВАННЯ УКРАЇНИ			
<i>Перший (бакалаврський) рівень СВО 126 Інформаційні системи і технології</i>	Кафедра Економічної кібернетики	Екзаменаційний білет № 24 з дисципліни «Аналіз і візуалізація даних»	Затверджую: Зав. кафедри <hr/> Проф. Жерліцин Д.М. ____.____.____ р.
1. Екзаменаційне питання (максимальна оцінка 10 балів за відповідь)			
Принципи побудови та аналізу зведених даних в Excel.			
2. Задача (максимальна оцінка 10 балів за розв'язання задачі)			
Стадо з 1500 голів протягом місяця годували спеціальним зерном з високим вмістом білка. Випадкова вибірка з 29 була зважена і набрала в середньому 6,7 фунтів. Якщо стандартне відхилення приросту ваги для всього стада становить 7,1, перевірте гіпотезу про те, що середній приріст ваги на бичка за місяць становив більше 5 фунтів.			
3. Тестові завдання різних типів (максимальна оцінка 10 балів за відповідь на всі тестові завдання)			

9. Методи навчання

При викладанні навчальної дисципліни використовуються такі *методи навчання*:

- M1. Лекція (дискусія, проблемна)
- M2. Лабораторна робота

- М3. Проблемне навчання
- М4. Он-лайн навчання

та методи контролю:

- МК1. Тестування
- МК2. Контрольне завдання
- МК3. Розрахункова робота
- МК4. Методи усного контроль
- МК5. Екзамен

10. Форми контролю

Кожна з форм контролю має особливості й залежить від мети, змісту та характеру навчання. У процесі навчання дисципліни використовуються наступні форми контролю:

- **Поточний контроль:** усне опитування (індивідуальне, фронтальне, групове), комп'ютерне тестування, виконання розрахункових завдань на комп'ютері згідно програми;
- **Підсумковий контроль:** іспит (тестування, теоретичне питання та розрахункова задача).

11. Розподіл балів, які отримують студенти.

Оцінювання знань студента відбувається за 100-бальною шкалою і переводиться в національні оцінки згідно з «Положення про екзамени та заліки у НУБіП України» (затверджено Вченою радою НУБіП України від 26.04.2023 р., протокол №10).

Поточний контроль		Рейтинг з навчальної роботи $R_{НР}$	Рейтинг з додаткової роботи $R_{ДР}$	Рейтинг штрафний $R_{ШТР}$	Підсумкова атестація (<i>icnum</i>)	Загальна кількість балів
Змістовий модуль 1	Змістовий модуль 2					
0-100	0-100	0-70	0-10	-	0-30	0-100

Для визначення рейтингу студента (слухача) із засвоєння дисципліни $R_{ДИС}$ (до 100 балів) одержаний рейтинг з атестації (до 30 балів) додається до рейтингу студента (слухача) з навчальної роботи $R_{НР}$ (до 70 балів):

$$R_{ДИС} = R_{НР} + R_{АТ}.$$

Загальний рейтинг здобувача вищої освіти із засвоєння дисципліни у балах переводиться у національної оцінки наступним чином:

Шкала оцінювання: національна та ECTS

Рейтинг здобувача вищої освіти, бали	Оцінка національна за результати складання екзаменів заліків	
	екзаменів	заліків
90-100	відмінно	зараховано
74-89	добре	
60-73	задовільно	
0-59	незадовільно	не зараховано

11. Методичне забезпечення

Електронний навчальний курс, розроблений на базі платформи LMS Moodle, розміщений на навчальному порталі НУБіП України за посиланням: <https://elearn.nubip.edu.ua/course/view.php?id=5316>.

12. Рекомендована література

Основна

1. Економетрика з R : навчальний посібник / А.В. Скрипник, Д.М. Жерліцин, Ю.О. Нам'ясенко. – Київ: ФОП Ямчинський О.В., 2020. – 248 с.
2. Статистичний аналіз та візуалізація даних: навчальний посібник / Д.М. Жерліцин, К.В. Наконечна, Л.В. Галаєва. – Київ: НУБіП. – 344 с.
3. Скрипник А.В. й ін. Теорія ймовірностей, імовірнісні процеси та математична статистика: навчальний посібник / Скрипник А.В., Галаєва Л.В., Коваль Т.В., Шульга Н.Г. – К.: ЦП "Компринт", 2017. – 320с.
4. Sarah Boslaugh (2012) Statistics in a Nutshell. Published by O'Reilly Media, Inc., 1005 Gravenstein Highway North, Sebastopol, 2012. 571 p.

Допоміжна

1. Глівенко С.В. Економічне прогнозування: Навчальний посібник / Глівенко С.В., Соколов М.О., Теліженко О.М. – 2-ге вид., перероб. та доп. – Суми: Видавництво «Університетська книга», 2001. – 207 с.
2. Демографічна та соціальна статистика / Доходи та умови життя. – К.: Державна служба статистики України, 2016. – 25 с.
3. Єріна А.М. Статистичне моделювання та прогнозування: Навч. Посібник / А.М. Єріна – К.: КНЕУ, 2014. – 340 с.
4. Кічор В. П. та ін. Економіко-статистичне моделювання і прогнозування: Навчальний посібник / В. П. Кічор, Р. В. Фещур, В. В. Козик, С. Н. Воробець, Н. Є. Семченко. – Львів: Видавництво Національного університету "Львівська політехніка", 2007. – 156 с.

5. Кулявець В.О. Прогнозування соціально-економічних процесів: Навчальний посібник / В.О. Кулявець – К.: Кондор, 2009. – 194 с.
6. Леоненко М.М. Теоретико-ймовірнісні та статистичні методи в економетриці та фінансовій математиці / М.М. Леоненко, Ю.С. Мішура, В.М. Пархоменко, М.Й. Ядренко. – К.: Інформтехніка, 1995. – 380 с.
7. Модернізація фінансових систем: методологія та інструменти управління / Ю.Г. Лисенко; Н.С. Педченко; В.М. Кравченко; В.В. Мандра та ін. / За ред. чл.кор. НАН України, д-ра екон. наук, проф. Лисенко Ю.Г.; д-ра екон. наук, доц. Жерлїцина Д.М. – Полтава, 2017. – 348 с.
8. Негрей М., Гнот Т. Аналітика з R: навчальний посібник / Негрей М., Гнот Т. Київ: ФОП Ямчинський О. В., 2020. – 236 с.
9. Оліскевич М.О. Основи економетрії часових рядів: навч. Посібник / М.О. Оліскевич. – Львів: ЛНУ імені І. Франка, 2009. – 327 с.
10. Скрипник А.В., Негрей М.В. Економетрика: навч. посібник. – Київ: КОМПРИНТ, 2017. – 272 с.
11. Теорія ймовірностей і математична статистика: підручник . Ч. 1 / М. А. Мартиненко, О. М. Нецадим, В. М. Сафонов. К. : , 2014. 287 с. [Електронний ресурс] - http://dspace.nu%D0%86RN_Ch1.pdf
http://bip.edu.ua:8080/jspui/bitstream/123456789/641/1/Martinenko_TEOR_JMOV
12. Щурик М.В. Статистика : навч. посіб. для вузів усіх рівнів акредитації / М.В. Щурик. – Львів : Магнолія, 2016. – 547 с.
13. James D. Hamilton. Time Series Analysis. Published by Princeton University Press, Chichester, West Sussex.

13. Інтернет ресурси

14. Державна служба статистики України [Електронний ресурс] / Режим доступу: <http://www.ukrstat.gov.ua/>
15. Підгорний А.З., Милашко О.Г. Система національних рахунків: навч. посіб. – Одеса, ОДЕУ, 2009 р. – 121 с. <http://dspace.oneu.edu.ua/jspui/bitstream/123456789/1053/1/Система%20національних%20рахунків.pdf>.
16. Державна комісія з регулювання ринків фінансових послуг України [Електронний ресурс] / Режим доступу: <http://www.dfp.gov.ua>
17. Національний банк України [Електронний ресурс] / Режим доступу: <http://www.bank.gov.ua>
18. ПФТС Фондова біржа [Електронний ресурс] / Режим доступу: <http://pfts.com.ua>
19. Украинская биржа – центр ликвидности интернет-трейдинга (акции, фьючерсы, опционы) [Электронный ресурс] / Режим доступа: <http://www.ux.ua>
20. Загальнодоступна інформаційна база даних Національної комісії з цінних паперів та фондового ринку про ринок цінних паперів [Електронний ресурс] / Режим доступу: <https://stockmarket.gov.ua/>