

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ
Волинський національний університет імені Лесі Українки
Факультет інформаційних технологій і математики
Кафедра комп'ютерних наук та кібербезпеки

СИЛАБУС
нормативного освітнього компонента
ІНТЕЛЕКТУАЛЬНИЙ АНАЛІЗ ДАНИХ
підготовки здобувачів освіти
першого (бакалаврського) рівня
спеціальності 122 Комп'ютерні науки
освітньо-професійної програми
Комп'ютерні науки та інформаційні технології

Силабус нормативного освітнього компонента “Інтелектуальний аналіз даних”
підготовки бакалаврів, галузі знань 12 – Інформаційні технології, спеціальності 122 –
Комп’ютерні науки, за освітньою програмою “Комп’ютерні науки та інформаційні
технології”

Розробник:

Мамчич Тетяна Іванівна, кандидат фізико-математичних наук, доцент, доцент кафедри
комп’ютерних наук та кібербезпеки

Погоджено

Гарант освітньо-професійної програми:



Гришанович Т. О.

**Силабус освітнього компонента затверджено на засіданні кафедри комп’ютерних наук та
кібербезпеки**

протокол № 2 від 29 вересня 2022 р.

Завідувач кафедри:



Гришанович Т. О.

ОПИС НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

Найменування показників	Галузь знань, спеціальність, освітня програма, освітній ступінь, ОКР	Характеристика навчальної дисципліни
		Нормативна
Денна форма навчання	Галузь знань 12 Інформаційні технології, спеціальність 122 Комп'ютерні науки, освітньо-професійна програма Комп'ютерні науки та інформаційні технології освітній рівень бакалавр.	Рік підготовки 2
Кількість годин /кредитів 120/4		Семестр 4
		Лекції 28 год.
		Лабораторні 34 год.
		Самостійна робота 50 год.
		Консультації 8 год.
ІНДЗ: <u>нема</u>		Форма контролю: екзамен

II Інформація про

викладача ІІІ: Мамчич Тетяна Іванівна;

Науковий ступінь: кандидат фізико математичних наук;

Вчене звання: доцент;

Посада: доцент кафедри комп'ютерних наук та кібербезпеки;

Контактна інформація: Mamchych.Tetyana@vnu.edu.ua

Дні занять: <http://194.44.187.20/cgi-bin/timetable.cgi>

III. Опис освітнього компонента

- Анотація курсу.** Силабус освітнього компонента складено відповідно до освітньо-професійної програми підготовки бакалаврів Комп'ютерні науки та інформаційні технології. Освітній компонент "Інтелектуальний аналіз даних" передбачає вивчення студентами основних методів роботи з числовими, порядковими та текстовими даними (Data Mining, Text Mining) разом з опануванням відповідних комп'ютерних технологій (мова програмування R). Перелік методів базується на сучасних статистичних підходах, залучаючи деякі алгоритми класифікацій, прогнозування та розпізнавання образів.

Предметом вивчення освітнього компонента є методи й технології аналізу даних.

- Мета навчальної дисципліни:** формування у слухачів знань, вмінь та навичок з аналізу даних, застосування відповідних інформаційних технологій для обробки даних.

3. Перелік компетентностей випускника

Загальні компетентності

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.

ЗК4. Здатність спілкуватися державною мовою як усно, так і письмово.

ЗК5. Здатність спілкуватися іноземною мовою.

ЗК6. Здатність вчитися й оволодівати сучасними знаннями.

ЗК7. Здатність до пошуку, оброблення та аналізу інформації з різних джерел.

ЗК8. Здатність генерувати нові ідеї (креативність).

ЗК10. Здатність бути критичним і самокритичним.

ЗК11. Здатність приймати обґрунтовані рішення.

ЗК12. Здатність оцінювати та забезпечувати якість виконуваних робіт.

ЗК13. Здатність діяти на основі етичних міркувань.

Спеціальні (фахові, предметні) компетентності

СК1. Здатність до математичного формулювання та досліджування неперервних та дискретних математичних моделей, обґрунтування вибору методів і підходів для розв'язування теоретичних і прикладних задач у галузі комп'ютерних наук, аналізу та інтерпретування.

СК2. Здатність до виявлення статистичних закономірностей недетермінованих явищ, застосування методів обчислювального інтелекту, зокрема статистичної, нейромережевої та нечіткої обробки даних, методів машинного навчання та генетичного програмування тощо.

СК6. Здатність до системного мислення, застосування методології системного аналізу для дослідження складних проблем різної природи, методів формалізації та розв'язування системних задач, що мають суперечливі цілі, невизначеності та ризики.

СК7. Здатність застосовувати теоретичні та практичні основи методології та технології моделювання для дослідження характеристик і поведінки складних об'єктів і систем, проводити обчислювальні експерименти з обробкою й аналізом результатів.

СК8. Здатність проектувати та розробляти програмне забезпечення із застосуванням різних парадигм програмування: узагальненого, об'єктно-орієнтованого, функціонального, логічного, з відповідними моделями, методами й алгоритмами обчислень, структурами даних і механізмами управління.

СК11. Здатність до інтелектуального аналізу даних на основі методів обчислювального інтелекту включно з великими та погано структурованими даними, їхньої оперативної обробки та візуалізації результатів аналізу впроцесі розв'язування прикладних задач.

Програмні результати навчання

ПРН1. Застосовувати знання основних форм і законів абстрактно-логічного мислення, основ методології наукового пізнання, форм і методів вилучення, аналізу, обробки та синтезу інформації в предметній області комп'ютерних наук.

ПРН4. Використовувати методи обчислювального інтелекту, машинного навчання, нейромережевої та нечіткої обробки даних, генетичного та еволюційного програмування для розв'язання задач розпізнавання, прогнозування, класифікації, ідентифікації об'єктів керування тощо.

ПРН12. Застосовувати методи та алгоритми обчислювального інтелекту та інтелектуального аналізу даних в задачах класифікації, прогнозування, кластерного аналізу, пошуку асоціативних правил з використанням програмних інструментів підтримки багатовимірної аналізу даних на основі технологій DataMining, TextMining, WebMining.

Структура освітнього компонента

Змістовий модуль 1. Класичні методи статистики.

Змістовий модуль 2. Сучасні комп'ютерні технології для аналізу даних.

Назви змістових модулів і тем	Усього	Лек.	Лабор.	Сам. роб.	Конс.	*Форма контролю / Бали
Змістовий модуль 1. Класичні методи статистики						
Тема 1. Описові статистики. Оцінювання параметрів.	6	2	2	2		Звіт по лаб. роботі / 2
Тема 2. Візуалізація даних.	9	2	2	4	1	Звіт по лаб. роботі / 2
Тема 3. Теорія оцінювання. Статистичне виведення.	11	2	4	4	1	Звіт по лаб. роботі / 4
Тема 4. Кореляційний аналіз. Регресійний аналіз.	13	4	4	4	1	Звіт по лаб. роботі / 4
Тема 5. Багатовимірні методи.	15	4	4	6	1	Звіт по лаб. роботі / 4
Разом за модулем 1	56	14	16	20	4	16
Змістовий модуль 2. Сучасні комп'ютерні технології для аналізу даних						
Тема 6. Аналіз та прогнозування динамічних рядів		2	4	4	1	Звіт по лаб. роботі / 4
Тема 7. Основи машинного навчання, нейронні мережі		2	2	8	1	Звіт по лаб. роботі / 2
Тема 8. Статистичний аналіз текстів (Text Mining)		2	4	6	1	Звіт по лаб. роботі / 6
Тема 9. Основи стохастичних методів аналізу (симуляції, стохаст. програмування).		4	4	6		Звіт по лаб. роботі / 6
Тема 10. Алгоритми роботи з порядковими та мережевими даними.		2	4	6	1	Звіт по лаб. роботі / 6
Разом за модулем 2	64	14	18	30	4	24
Види підсумкових робіт						
Модульна контрольна робота 1						30
Модульна контрольна робота 2						30
Всього годин/ Балів 120/ 100	120	28	34	50	8	100

Завдання для самостійного опрацювання

№ з/п	Тема	Кількість годин
1	Підготовка до лабораторних робіт.	20
2	Опрацювання лекційного матеріалу. Підготовка до контрольних робіт.	20
3	Систематизація здобутих знань перед екзаменом	10
	Разом	50

IV. Політика оцінювання

Політика щодо академічної доброчесності. Академічна доброчесність базується на засудженні практик списування (виконання письмових робіт із залученням зовнішніх джерел інформації, крім дозволених для використання), плагіату (відтворення опублікованих текстів інших авторів без зазначення авторства), фабрикації (вигадування даних чи фактів, що використовуються в освітньому процесі). У разі порушення здобувачем вищої освіти академічної доброчесності (списування, плагіат, фабрикація), робота оцінюється незадовільно та має бути виконана повторно, а результати раніше зданих робіт анулюються і виконуються повторно у порядку визначеному викладачем. При цьому викладач залишає за собою право змінити завдання.

Комунікаційна політика. Здобувачі вищої освіти повинні мати активовану університетську пошту. Усі письмові запитання до викладачів стосовно курсу мають надсилатися на університетську електронну пошту, можливе інше (додаткове) джерело комунікації, визначене викладачем для більш оперативного зв'язку зі студентами.

Політика щодо перескладання. Роботи, які здаються із порушенням термінів без поважних причин оцінюються на нижчу оцінку. Перескладання модулів відбувається із дозволу лектора за наявності поважних причин (наприклад, лікарняний).

Політика щодо оскарження оцінювання. Політика щодо оскарження оцінки.

Якщо здобувач вищої освіти не згоден з оцінюванням його знань він може опротестувати виставлену викладачем оцінку у встановленому порядку згідно «Положення про порядок і процедури вирішення конфліктних ситуацій у Волинському національному університеті імені Лесі Українки»

Політика щодо відвідування занять. Для здобувачів вищої освіти денної форми відвідування занять є обов'язковим. Поважними причинами для неявки на заняття є хвороба, академічна мобільність, які необхідно підтверджувати відповідними документами. Про відсутність на занятті та причини відсутності здобувач вищої освіти має повідомити викладача або особисто, або через старосту. За об'єктивних причин навчання може відбуватись в он-лайн формі за погодженням з керівником курсу та деканом факультету.

Визнання результатів навчання, отриманих у формальній, неформальній освіті. Під час вивчення освітнього компонента можливе визнання результатів навчання отриманих у формальній, неформальній та/або інформальній освіті. Порядок визнання результатів навчання для здобувачів вищої освіти, набутих у: формальній освіті (академічна мобільність студентів на території України чи поза її межами, для студентів, які переводяться, поновлюються з інших ЗВО (вітчизняних чи іноземних); неформальній та/або інформальній освіті здійснюється згідно «ПОЛОЖЕННЯ про визнання результатів навчання, отриманих у формальній, неформальній та/або інформальній освіті у Волинському національному університеті імені Лесі Українки» (<https://cutt.ly/yNUt5Y4>).

При оцінюванні може бути врахований курс "Аналіз даних та статистичне виведення з програмою R" (30 балів), "Візуалізація даних"(10 балів) на платформі відкритих курсів Prometheus.

Підсумковий контроль

Форма контролю – семестровий екзамен. Оцінювання здійснюється за 100-бальною шкалою. Оцінка включає в себе поточний контроль (нараховується за якісне виконання лабораторних робіт) та підсумковий модульний контроль (нараховується за виконання модульних контрольних робіт та модульних тестових робіт). Максимальна кількість балів, яку може отримати студент під час поточного оцінювання за семестр – 40 балів. Підсумковий модульний контроль за семестр включає в себе оцінки за всі модульні контрольні роботи, тестові завдання і складає 60 балів.

Якщо за результатами семестру накопичено не менше 75 балів і студент погоджується із цим результатом, то оцінка за семестр може виставлятися без складання іспиту. В іншому випадку студент складає іспит; максимальна кількість балів, яку можна отримати на іспиті – 60 балів при цьому бали за підсумковий модульний контроль анулюються.

Екзамен проходить в письмовій формі. Оцінка за семестр у випадку складання іспиту є сумою балів поточного контролю та балів, отриманих під час іспиту.

Питання, які виносяться на екзамен

1. Типи показників для статистичного аналізу. Приклади.
2. Статистичний розподіл даних, способи задання розподілу.
3. Візуалізація статистичного розподілу. Гістограма, кумулята, графік, полігон частот, кругова діаграма, діаграма “ящик з вусами”.
4. Числові характеристики рівня та розкиду статистичних показників: середнє, мода, медіана, дисперсія, середнє квадратичне відхилення, стандартна похибка, квантілі, квартилі, децилі, децильний коефіцієнт, коефіцієнт варіації, розмах варіації.
5. Точкові та інтервальні оцінки. Зв'язок точності та надійності довірчих інтервалів.
6. Основні поняття статистичного виведення: перевірка статистичних гіпотез, помилки I і II роду, рівень значущості, потужність критерію.
7. Призначення критеріїв Пірсона χ^2 , Стюдента, Фішера, Шапіро, Колмогорова-Смірнова, ANOVA, MANOVA.
8. Кореляційний аналіз. Коефіцієнти зв'язку для двох показників.
9. Парна та множинна регресія. Метод найменших квадратів.
10. Аналіз та прогнозування динамічних рядів, у тому числі із “сезонними змінами”.
11. Кластерний аналіз.
12. Факторний аналіз.
13. Дискримінантний аналіз.
14. Поняття про машинне навчання: розпізнавання образів, регресія та класифікація, алгоритми опорних векторів (SVM). Нейронні мережі (NN). Статистичні процедури Bootstrap.
15. Поняття про стохастичні обчислювальні алгоритми: симуляції, метод Монте-Карло, метод випадкового спуску для оптимізації, мурашиний алгоритм (Ant Colony Optimization), бджолиний алгоритм (Artificial Bee Colony Algorithm), генетичні алгоритми.
16. Аналіз текстових даних (Text Mining), контент-аналіз заданого корпусу даних. Сучасні можливості програми R для пошуку та аналізу текстових даних.
17. Деякі методи аналізу порядкових даних: знаходження колективних рішень (social choice), метрика Кемені-Снелла.
18. Моделі графів для мережевих даних. Збалансованість у моделях знакових графів. Міра статусу Харарі в моделях орієнтованих графів для ієрархічних структур. Моделі ланцюгів Маркова для інформаційних мереж.
19. Обчислювальні можливості програми R: обчислення в інтерактивному режимі, вектори та матриці, дата фрейми, створення та виконання скриптів, введення та виведення даних, імпорт та експорт.
20. Використання пакетів у програмі R. Статистичні функції для описової статистики, візуалізації даних, перевірки гіпотез, багатовимірних методів, Text Mining, SVM, NN.

Екзамен проходить письмово у вигляді складання звіту з відповідями на теоретичні питання та з результатами виконання практичних завдань за допомогою відповідних комп'ютерних програм (копіювання виконаних команд).

На обчислення, підготовку відповідей та оформлення звіту відводиться 100 хвилин.

Екзаменаційний білет містить 5 теоретичних питань (це можуть також бути завдання на обчислення без комп'ютерних програм) та 5 практичних завдань на обчислення за допомогою програми R. Для обчислень кожному студенту пропонується індивідуальний набір даних.

Приклад екзаменаційного білета:

Всі питання по 6 балів.

Теоретичні питання (короткі письмові відповіді):

1. Для заданого набору дискретних даних знайти середнє та медіану без використання комп'ютерних програм.
2. Метод найменших квадратів. Лінія регресії.
3. Метод опорних векторів для класифікації багатовимірних даних.
4. Факторний аналіз: призначення та умови застосовності.
5. Обчислити результат колективного вибору для набору індивідуальних профілів (порядкових даних).

Практичні завдання (виконуються з використанням мови програмування R):

6. Для заданого набору даних обчислити середнє та медіану.
7. Для заданого набору даних побудувати гістограму частот, додати графік щільності.
8. Для заданого набору текстових даних знайти матрицю частот для вказаних лексем (термінів),
9. Обчислити коефіцієнт кореляції для двох заданих документів відносно вживання заданого набору термінів (дані із завд. 8).
10. Для заданого набору даних побудувати лінійну регресію.

Шкала оцінювання (національна та ECTS)

Оцінка в балах	Лінгвістична оцінка	Оцінка за шкалою ECTS	
		оцінка	пояснення
90–100	Відмінно	A	відмінне виконання
82–89	Дуже добре	B	вище середнього рівня
75–81	Добре	C	загалом хороша робота
67–74	Задовільно	D	непогано
60–66	Достатньо	E	виконання відповідає мінімальним критеріям
1–59	Незадовільно	Fx	Необхідне перескладання

Рекомендована література та інтернет-ресурси

1. Майборода Р.Є. *"Комп'ютерна статистика"*. ВПЦ "Київський університет", 2019. 589 с.
<http://probability.univ.kiev.ua/userfiles/mre/cscolor.pdf>
2. Майборода Р. Багатовимірна дескриптивна статистика. Методичні рекомендації по курсу. КНУ ім. Тараса Шевченка, 2020.
<https://probability.knu.ua/userfiles/mre/descrtasks.pdf>
3. Ланде Д.В., Субач І.Ю., Бояринова Ю.Є. Основи теорії і практики інтелектуального аналізу даних у сфері кібербезпеки: навчальний посібник. К.: ІСЗІ КПІ ім. Ігоря Сікорського, 2018. 300 с.
4. TextMining with R : <http://www.slideshare.net/whitish/textmining-with-r>
5. A. Agresti. Statistical methods for social sciences. Boston: Pearson. 2018.
6. The R Project for Statistical Computing <https://www.r-project.org/>
7. Сергеев-Горчинський О.О., Іщенко Г.В. Інтелектуальний аналіз даних: комп'ютерний практикум. КПІ ім. Ігоря Сікорського, 2018. 73 с.
https://ela.kpi.ua/bitstream/123456789/24971/1/Komp_prakt.pdf

MINISTRY OF EDUCATION AND SCIENCE OF UKRAINE
Lesya Ukrainka Volyn National University
Faculty of Information Technologies and Mathematics
Department of Computer Science and Cyber Security

SYLLABUS
of the normative educational component
INTELLIGENT DATA ANALYSIS
for the first (bachelor) level of higher education,
specialty 122 Computer science
educational program
Computer science and information technologies

Syllabus of the normative educational component "Intelligent data analysis" of bachelor's training, field of knowledge 12 - Information technologies, specialty 122 - Computer science, educational program "Computer science and information technology"

Developer:

Tetyana Mamchych, PhD on Mathematics, Associate Professor, Associate Professor of the Department of Computer Sciences and Cyber Security

Agreed



Guarantor of the educational program Hryshanovych T. O.

The syllabus of the educational component was approved at the meeting of the Department of Computer Sciences and Cyber Security
protocol No. 2 dated 09/29/2022

Head of the department



Hryshanovych T.O.

I DESCRIPTION OF THE EDUCATIONAL DISCIPLINE

Indicators	Field of knowledge, specialty, educational program, educational level	Characteristics of the subject
		Normative
Full-time	Field of knowledge 12 Information technologies, specialty 122 Computer science, educational program Computer science and information technology	Year 2
Number of hours/кредитів 120/4		Semester 4
		Lectures: 28 hrs
		Laboratory works: 34 hrs
		Independent work: 50 hrs
		Consultations: 8 hrs
Individual task: none		Form of evaluating: exam

II Information about the teacher:

Tetyana Ivanivna Mamchych;

Academic degree: PhD in Mathematics;

Academic title: Associate professor;

Position: Associate Professor of the Department of Computer Science and Cyber Security;

Contact information: Mamchych.Tetyana@vnu.edu.ua

Class days: <http://194.44.187.20/cgi-bin/timetable.cgi>

III. Description of the educational component

Abstract of the course. The syllabus of the educational component is compiled in accordance with the educational and professional program of Bachelors of Computer Science and Information Technologies. The educational component "Intelligent data analysis" consists of the basic methods of processing numerical, ordinal and text data (Data Mining, Text Mining) along with mastering the relevant computer technologies (R program). The list of methods is based on modern statistical approaches, involving some classification, forecasting and pattern recognition algorithms.

The subject of study of the educational component is methods and technologies of data analysis.

The purpose of the educational discipline: formation of students' knowledge, skills and abilities in data analysis, application of appropriate information technologies for data processing.

The list of competences of a graduate

General competences (GC)

- GC1.** Ability to abstract thinking, analysis and synthesis.
- GC2.** Ability to apply knowledge in practical cases.
- GC3.** Knowledge and understanding of the subject area and understanding a professional activity.
- GC4.** Ability to communicate in the state language both orally, and in writing.
- GC5.** Ability to communicate in a foreign language.
- GC6.** Ability to learn and master modern knowledge.
- GC7.** Ability to search, process and analyze information from various sources.
- GC8.** The ability to generate new ideas (creativity).
- GC10.** The ability to be critical and self-critical.
- GC11.** Ability to make reasoned decisions.
- GC12.** The ability to evaluate and provide the quality of the work performed.
- GC13.** The ability to act on the basis of ethical considerations.

Special (professional) competences (SC)

- SC1.** Ability to mathematically formulate and investigate continuous and discrete mathematical models, justify the choice of methods and approaches for solving theoretical and applied problems in the field of computer science, analysis and interpretation.
- SC2.** The ability to identify statistical regularities of non-deterministic phenomena, use methods of computational intelligence, in particular statistical, neural network and fuzzy data processing, methods of machine learning and genetic programming, etc.
- SC6.** Ability to system thinking, application methodology of system analysis for the study of complex problems of various nature, methods of formalization and solving system problems that have contradictory objectives, uncertainties and risks.
- SC7.** Ability to apply theoretical and practical basics of modeling methodology and technology for study of the characteristics and behavior of complex objects and systems, conduct computational experiments with processing and analysis of results.
- SC8.** Ability to design and develop software provision using different paradigms programming: generalized, object-oriented, functional, logical, with appropriate models, calculation methods and algorithms, data structures and management mechanisms.
- SC11.** Ability to intelligently analyze data based on methods of computational intelligence, including large and poorly structured data, data processing and visualization while solving practical problems.

Expected learning outcomes (ELO).

- ELO1.** Apply knowledge of the basic forms and laws of abstract and logical thinking, the basics of the methodology of scientific knowledge, the forms and methods of extracting, analyzing, processing and synthesizing information in the subject area of computer science.
- ELO4.** Use methods of computational intelligence, machine learning, neural network and fuzzy data processing, genetic and evolutionary programming for solving tasks of recognition, forecasting, classification, identification of control objects, etc.
- ELO12.** Apply methods and algorithms of computational intelligence and intelligent data analysis in tasks of classification, forecasting, cluster analysis, search for associative rules using software tools to support multidimensional data analysis based on DataMining, TextMining, WebMining technologies.

IV The structure of the Educational component

Content module 1. Classical methods of statistics.

Content module 2. Modern computer technologies for data analysis.

Topics	Total (hrs)	Lectures (hrs)	Lab. Work (hrs)	Selflearning (hrs)	Consult. (hrs)	*Evaluation / Points
Module 1. Classical methods of statistic						
1. Descriptive statistics. Estimation of parameters.		2	2	2		10
2. Data visualization.		2	2	4	1	10
3. Statistical inference.		2	4	4	1	10
4. Correlation. Regression.		4	4	4	1	10
5. Multivariate statistical methods.		4	4	6	1	10
6. Total for Module 1	56	14	16	20	4	50
Module 2. Modern computer technologies for data analysis						
6. Time series analysis and forecasting		2	4	4	1	10
7. Introduction to machine learning and neural nets.		2	2	8	1	10
8. Text Mining		2	4	6	1	10
9. Introduction to stochastic methods and simulations.		4	4	6		10
10. Some models and algorithms for ordered data and nets.		2	4	6	1	10
Total for Module 2	64	14	18	30	4	50
Final evaluation	Current evaluation (for the semester) (40 points) Module 1 (30 points)+Module 2 (30 points) or Exam (60 points)					
Total (Hours/ Points) 120/ 100	120	28	34	50	8	100

Tasks for individual work

№	Topics	Time (hrs)
1.	Laboratory work preparation	20
2.	Refreshing material from lectures. Module test preparation.	20
3.	Exam preparation	10
4.	Total	50

V Evaluation policy

Academic Integrity Policy

Academic integrity is based on condemning the practices of plagiarism (making written works involving external sources of information other than those permitted for use), plagiarism (reproducing published texts of other authors without indicating authorship), fabrication (inventing data or facts used in the educational process). In the case of a violation of academic integrity by a student (copying, plagiarism, fabrication), the work is evaluated as unsatisfactory and must be re-done, and the results of previously passed works are canceled and re-done in the order determined by the teacher. At the same time, the teacher reserves the right to change the assignment.

Communication policy

Applicants for higher education must have an activated university mail. All written questions to the lecturers about the course should be sent to the university e-mail, possibly another (additional) source of communication, determined by the lecturer for more prompt communication with students.

Reorder Policy

Works that are submitted late without good reason are evaluated at a lower grade. Modules can be rearranged with the permission of the lecturer if there are good reasons (for example, sick leave).

Assessment Appeals Policy

If the student does not agree with the assessment of his knowledge, he can protest the grade issued by the teacher in accordance with the established procedure according to the "Regulations on the procedure and procedures for resolving conflict situations at Lesya Ukrainka Volyn National University"

Class attendance policy

For full-time higher education students, attending classes is mandatory. Important reasons for not attending classes are illness, academic mobility, which must be confirmed with relevant documents. The student of higher education must inform the teacher about the absence from the class and the reasons for the absence either personally or through the headmaster. For objective reasons, training can take place online upon agreement with the course leader and the dean of the faculty.

Recognition of learning outcomes obtained in formal, non-formal education

During the study of the educational component, it is possible to recognize the learning outcomes obtained in formal, informal and/or informal education. The procedure for recognizing the results of higher education for students of higher education acquired in: formal education (academic mobility of students on the territory of Ukraine or outside of Ukraine, for students who are transferring, renewing from other higher education institutions (domestic or foreign); informal and/or informal education is carried out in accordance "REGULATION on the recognition of learning outcomes obtained in formal, informal and/or informal education at Lesya Ukrainka Volyn National University"

(<https://cutt.ly/yNUt5Y4>). The course "Data analysis and statistical inference with the R program" (30 points), "Data visualization" (10 points) on the Prometheus open course platform can be taken into account during the evaluation.

Final evaluation

The form of control is a semester's exam. Evaluation is carried out on a 100-point scale. The assessment includes current control (accrued for the quality performance of laboratory works) and final modular control (accrued for the performance of modular control works and modular test works). The maximum number of points that a student can receive during the current assessment for the semester is 40 points. The final module control for the semester includes grades for all module control works, test tasks and is 60 points.

If at least 75 points are accumulated according to the results of the semester and the student agrees with this result, then the grade for the semester can be issued without taking the exam. Otherwise, the student takes the exam; the maximum number of points that can be obtained on the exam is 60 points, while the points for the final module control are canceled.

The exam is held in written form. The grade for the semester in the case of passing the exam is the sum of the points of the current control and the points obtained during the exam.

Exam questions

1. Types of variables(indicators) for statistical analysis. Examples.
2. Statistical distribution of data.
3. Visualization of statistical distribution. Histogram, cumulative chart, graph, frequency polygon, pie chart, box-whisker chart.
4. Numerical characteristics of the level and spread of statistical indicators: mean, mode, median, variance, standard deviation, standard error, quantiles, quartiles, deciles, decile coefficient, coefficient of variation, range of variation.
5. Point and interval estimations. Dependence of accuracy and reliability of confidence intervals.
6. Basic concepts of statistical inference: testing of statistical hypotheses, errors of the type I and II, significance level, power of a test.
7. Tests of Pearson χ^2 , Student, Fisher, Shapiro, Kolmogorov-Smirnov, ANOVA, MANOVA methods.
8. Correlation analysis. Correlation coefficients for two indicators.
9. Pair and multiple regression. The least square method.
10. Analysis and forecasting of dynamic series, including with "seasonal changes".
11. Cluster analysis.
12. Factor analysis.
13. Discriminant analysis.
14. Concepts of machine learning: pattern recognition, regression and classification, support vector algorithms (SVM). Neural networks (NN). Bootstrap statistical procedures.
15. Concepts of stochastic computing algorithms: simulations, Monte Carlo method, random descent method for optimization, ant algorithm (Ant Colony Optimization), bee algorithm (Artificial Bee Colony Algorithm), genetic algorithms.
16. Text data analysis (Text Mining), content analysis of a given corpus of data. Modern capabilities of the R program for searching and analyzing textual data.
17. Some methods of analyzing ordinal data: finding collective decisions (social choice), the Kemeny-Snell metric.

18. Graph models for network data. Balance in models of sign graphs. Harari status measure in directed graph models for hierarchical structures. Markov chain models for information networks.
19. Computing with R program: calculations in interactive mode, vectors and matrices, data frame, writing and running of scripts, input and output of data, import and export.
20. Using packages in R. Statistical functions for descriptive statistics, data visualization, hypothesis testing, multivariate methods, Text Mining, SVM, NN.

The exam is conducted in a written form. The written report should contain answers for theoretical questions and the results of practical tasks (copying executed commands from the console).

100 minutes are allocated for calculations, preparation of answers and preparation of the report.

The exam's list of questions contains 5 theoretical questions (these can also be tasks for calculations without computer programs) and 5 practical tasks for calculations using the R program. For calculations each student is offered an individual set of data.

An example of an examination ticket:

All questions are worth 6 points.

Theoretical questions (short written answers):

1. For a given set of discrete data, find the mean and median without using computer programs.
2. The method of least squares. Regression line.
3. The method of support vectors for the classification of multidimensional data.
4. Factor analysis: purpose and conditions of applicability.
5. Calculate the result of collective choice for a set of individual profiles (ordinal data).

Practical tasks (performed with the R program):

6. For a given set of data, calculate the mean and median.
7. For the given set of data create a frequency histogram, add a density line.
8. For a given set of textual data, find the frequency matrix for the indicated lexemes (terms).
9. Calculate the coefficients of correlation for two given documents regarding the use of a given set of terms (data from task 8).
10. Construct a linear regression for the given data set.

Grading scale (national and ECTS)

A scale for evaluating the knowledge of education seekers on educational components, where the form of control is an exam

Score in points	Linguistic score	ECTS score	
			Explanation
90–100	Excellent	A	Outstanding performance without errors
82–89	Good	B	Above the average standard but with minor errors
75–81	Good	C	Generally sound work with some errors
67–74	Satisfactorily	D	Fair but with significant shortcomings
60–66	Satisfactorily	E	Performance meets the minimum criteria
1–59	Unsatisfactorily	Fx	Fail. Re-examination is required

Recommended literature and Internet resources

1. Maiboroda R.E. "Computer statistics". VOC "Kyiv University", 2019. 589 p. (in ukr.)
<http://probability.univ.kiev.ua/userfiles/mre/cscolor.pdf>
2. Maiboroda R. Multivariate descriptive statistics. Methodical recommendations for the course. KNU named after Taras Shevchenko, 2020. (in ukr.)
<https://probability.knu.ua/userfiles/mre/descrtasks.pdf>
3. Lande D.V., Subach I.Yu., Boyarynova Yu.E. Fundamentals of the theory and practice of intelligent data analysis in the field of cyber security: a study guide. K.: ISZZI KPI named after Igor Sikorskyi, 2018. 300 p. (in ukr.)
4. TextMining with R: <http://www.slideshare.net/whitish/textmining-with-r>
5. A. Agresti. Statistical methods for social sciences. Boston: Pearson. 2018.
6. The R Project for Statistical Computing <https://www.r-project.org/>
7. Sergeyev-Gorchynskyi O.O., Ishchenko G.V. Intelligent data analysis: computer workshop. KPI named after Igor Sikorskyi, 2018. 73 p. (in ukr.)
https://ela.kpi.ua/bitstream/123456789/24971/1/Komp_prakt.pdf
8. Trevor Hastie, Robert Tibshirani, and Jerome Friedman. The Elements of Statistical Learning.
<https://hastie.su.domains/ElemStatLearn/>
9. Data Science Full Course - Learn Data Science in 10 Hours | Data Science For Beginners | Edureka
<https://www.youtube.com/watch?v=-ETQ97mXXF0>
10. Statistics - A Full University Course on Data Science Basics
<https://www.youtube.com/watch?v=xxpc-HPKN28>
11. T. Mamchych. Intelligent data analysis. Part 1. Applied statistics. Lecture Notes. Lesya Ukrainka Volyn National University. 2022. (in ukr.)