

# Чернівецький національний університет імені Юрія Федьковича

(повне найменування закладу вищої освіти)

## Факультет математики та інформатики

(назва інституту/факультету)

Кафедра математичного моделювання

(назва кафедри)

## СИЛАБУС

навчальної дисципліни

### Аналіз і візуалізація даних в R

вибіркова

(вказати: обов'язкова)

Освітньо-професійна програма "Інформаційні технології та управління проектами"

(назва програми)

Спеціальність 122 – Комп'ютерні науки

(вказати: код, назва)

Галузь знань 12 - Інформаційні технології

(вказати: шифр, назва)

Рівень вищої освіти перший (бакалаврський)

(вказати: перший (бакалаврський)/другий (магістерський)/третій (освітньо-науковий))

факультет математики та інформатики

(назва факультету/інституту, на якому здійснюється підготовка фахівців за вказаною освітньо-професійною програмою)

Мова навчання українська

(вказати: на яких мовах читається дисципліна)

Розробники: канд. ф.-м.н., доц. Дорошенко І.В.

(вказати авторів (викладач (ів)), їхні посади, наукові ступені, вчені звання)

Профайл викладача (-ів) <https://mathmod.chnu.edu.ua/pro-nas/spivrobotnyky/doroshenko-iryna-viktorivna/>

Контактний тел.

**0504340655**

E-mail:

**i.doroshenko@chnu.edu.ua**

Сторінка курсу в Moodle <https://moodle.chnu.edu.ua/course/view.php?id=4682>

Консультації

## **1. Анотація дисципліни (призначення навчальної дисципліни).**

Призначення дисципліни – формування у здобувачів освіти уміння застосовувати сучасні комп'ютерні технології до постановки, аналізу та розв'язання основних задач статистичного аналізу даних із застосуванням середовища R.

**2. Мета навчальної дисципліни:** вивчення основ роботи у статистичному програмному середовищі R та оволодіння методами розв'язання основних задач математичної статистики у цьому середовищі, візуалізація даних за допомогою відомих бібліотек мови R.

**Основними завданнями:** формування практичних умінь у сфері статистики; основних принципів роботи з R; основних графічних оболонок; основних структурних одиниць мови програмування R та операцій над ними; основних елементів програмування: циклів, розгалужень, функцій; основних методів графічного відображення даних та результатів їх обробки за допомогою спеціального пакету ggplot2; методів роботи засобами мови R з основними ймовірнісними розподілами; методів розв'язання засобами мови R задач описової статистики; методів перевірки засобами мови R основних статистичних гіпотез про вид розподілу вибірки; методів знаходження коефіцієнтів простої лінійної регресії; методів діагностики моделей регресії.

**3. Пререквізити.** Теорія ймовірностей та математична статистика, програмування.

## **4. Результати навчання:**

Формування компетентностей, а саме

***Загальні та фахові компетентності:***

ЗК1. Здатність до абстрактного мислення, аналізу та синтезу.

ЗК2. Здатність застосовувати знання у практичних ситуаціях.

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

ЗК6. Здатність вчитися й оволодівати сучасними знаннями.

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

ЗК8. Здатність генерувати нові ідеї (креативність).

ЗК9. Здатність працювати в команді.

ЗК11. Здатність приймати обґрунтовані рішення.

ФК1. 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.

ФК2. 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.

### **Програмні результати навчання:**

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

ПРН3. Використовувати знання закономірностей випадкових явищ, їх властивостей та операцій над ними, моделей випадкових процесів та сучасних програмних середовищ для розв'язування задач статистичної обробки даних і побудови прогнозних моделей.

ПРН8. Використовувати методологію системного аналізу об'єктів, процесів і систем для задач аналізу, прогнозування, управління та проектування динамічних процесів в макроекономічних, технічних, технологічних і фінансових об'єктах.

## **5. Опис навчальної дисципліни**

### **5.1. Дидактична карта навчальної дисципліни**

Назви змістових модулів і тем	Кількість годин											
	денна форма						Заочна форма					
	усього	у тому числі					усього	у тому числі				
		л	п	лаб	інд	с.р.		л	п	лаб	інд	с.р.
1	2	3	4	5	6	7	8	9	10	11	12	13
<b>Змістовний модуль 1. Зведення, класифікація та групування первинних даних. Статистичні показники</b>												
Тема 1. Основи роботи в системі R	8	2	-	1		5						
Тема 2. Основи математичної статистики в R	21	2	2	2		15						
Тема 3. Графічні можливості R	8	1	1	1		5						
Тема 4. Робота з розподілами	5	1	2	2		-						
Тема 5. Квартилі, децилі, процентилі. Асиметрія і ексцес	21	2	2	2		15						
Разом за змістовим модулем 1	63	8	7	8		40						

<b>Змістовий модуль 2. Основні види статистичного аналізу в R</b>												
Тема 1. Перевірка статистичних гіпотез в R	9	1	2	1		5						
Тема 2. Елементи дисперсійного аналізу в R	16	2	2	2		10						
Тема 3. Елементи кореляційно аналізу в R	16	2	2	2		10						
Тема 4. Основи регресійного аналізу в R	16	2	2	2		10						
<b>Разом за змістовим модулем 2</b>	<b>57</b>	<b>7</b>	<b>8</b>	<b>7</b>		<b>35</b>						
<b>Усього годин</b>	<b>120</b>	<b>15</b>	<b>15</b>	<b>15</b>		<b>75</b>						

## 5.2. Самостійна робота студента

№ з/п	Назва теми	Кількість годин
1	Основи роботи з R та RStudio. Встановлення системи, RStudio. Робота зі скриптами. Встановлення пакетів. Використання R в якості калькулятора. Основні арифметичні операції над векторами.	10
2	Найпростіші об'єкти R. Матриці. Масиви. Фактори.	5
3	Складні структурні одиниці мови R. Списки. Фрейми. Функції групи apply.	10
4	Візуалізація даних. Функції qplot, ggplot. Різноманітні geoms.	10
5	Описова статистика. Гістограма. P-P і Q-Q діаграми. Обчислення числових характеристик вибірок.	10
6	Перевірка гіпотез про вид розподілу. Критерії перевірки гіпотези про розподіл вибірки: Колмогорова-Смірнова, критерій .	10
7	Елементи кореляційно-регресійного аналізу	10
8	Знаходження коефіцієнтів кореляції Фехнера, Спірмена та Кендала	10

## **6. Освітні технології, методи навчання і викладання навчальної дисципліни**

У викладання курсу використовуються такі освітні технології:

- Лекції та їх презентації.
  - Онлайн-лекції.
  - Лабораторні заняття.
  - Практичні заняття
  - Групова робота, коли студенти розв'язують практичні завдання.
  - Онлайн-тести та опитування: Використання системи MOODLE
- Електронні підручники і ресурси репозитарію ЧНУ

### **Методи навчання:**

МН 1 - лекція-візуалізація;

МН 8 – робота з тестами;

МН 9 – робота в групах;

МН 12 – дистанційне навчання з використанням відповідних онлайн-платформ

## **7. Контроль та оцінювання результатів навчальних досягнень студентів з навчальної дисципліни**

### **Види та форми контролю**

1. Поточний (захист лабораторних робіт, опитування теоретичного матеріалу)
2. Модульний (тестування, виконання завдань)
3. Підсумковий (залік)

### **Засоби оцінювання**

Засобами оцінювання та демонстрування результатів навчання можуть бути:

- перевірка викладачем та захист студентами письмових звітів про виконання кожної практичної роботи,
- експрес-опитування,
- тестові завдання.

Система оцінювання рівня навчальних досягнень ґрунтується на принципах ECTS та є накопичувальною. Протягом семестру студенти виконують 7 лабораторних робіт та 3 практичних робіт. Кожна лабораторна та практична роботи оцінюється кількістю балів за таблицею, наведеною нижче.

Виконуючи завдання лабораторної та практичної роботи, студент повинен оформити і завантажити для подальшої перевірки на сайт електронного навчання звіт разом із працездатними файлами програмної реалізації завдань ЛР та ПР (правила оформлення наведені на сторінці навчальної дисципліни на сайті).

50% балів, відведених на оцінювання ЛР, студент отримує за працюючий програмний продукт, в якому реалізовано всі завдання ЛР та оформлений звіт. Решта 50% балів виставляється після захисту студентом виконаного звіту. На захисті звіту з ЛР студент має відповісти на питання щодо постановки задачі та розробленого ним алгоритму реалізації кожного із завдань ЛР. При відповіді на теоретичні питання та питання щодо програмної реалізації алгоритму у випадку неістотної помилки знімається 10-20% балів, а у випадку істотної 20-40% балів, якщо ж студент не опанував теоретичний матеріал, плутається в означеннях, наводить логічно невірні твердження, то знімається до 50% балів від усієї суми балів за ЛР.

50% балів, відведених на оцінювання ПР, студент отримує за правильно розв'язанні завдання та оформлений звіт. Решта 50% балів виставляється після захисту студентом виконаного звіту. На захисті звіту з ПР студент має відповісти на питання щодо постановки задачі та алгоритму розв'язання завдань ПР. При відповіді на теоретичні питання у випадку неістотної помилки знімається 10-20% балів, а у випадку істотної 20-40% балів, якщо ж

студент не опанував теоретичний матеріал, плутається в означеннях, наводить логічно невірні твердження, то знімається до 50% балів від усієї суми балів за ПР.

Максимальна кількість, яку можна набрати на підсумковому модулі (тестування) – 40 балів.

Підсумкова оцінка виставляється за результатами суми балів, набраних на змістових модулях під час семестру та підсумковому модулі згідно з нижче наведеною таблицею.

**Критерії оцінювання результатів навчання з навчальної дисципліни  
Розподіл балів, які отримують студенти**

Поточне оцінювання (аудиторна та самостійна робота)									Кількість балів (залік)	Сумарна к-ть балів
Змістовий модуль №1					Змістовий модуль № 2					
T1	T2	T3	T4	T5	T1	T2	T3	T4	40	100
5	10	5	5	5	5	5	10	0		

**Дедлайни та перескладання.** Роботи, які здаються із порушенням термінів без поважних причин, оцінюються на нижчу оцінку. Перескладання модулів відбувається з дозволу деканату за наявності поважних причин (наприклад, лікарняний, участь у конференції, студентській олімпіаді).

**Шкала оцінювання: національна та ЄКТС**

Оцінка за національною шкалою	Оцінка за шкалою ECTS	
	Оцінка (бали)	Пояснення за розширеною шкалою
<b>Зараховано</b>	A (90-100)	відмінно
	B (80-89)	дуже добре
	C (70-79)	добре
	D (60-69)	задовільно
	E (50-59)	достатньо
<b>Незараховано</b>	FX (35-49)	(незадовільно) з можливістю повторного складання
	F (1-34)	(незадовільно) з обов'язковим повторним курсом

**7. Політика освітнього процесу**

Здобувач зобов'язаний своєчасно та якісно виконувати всі отримані завдання; за необхідністю з метою з'ясування всіх незрозумілих під час самостійної та індивідуальної роботи питань, відвідувати консультації викладача.

Студенти мають дотримуватись правил академічної доброчесності відповідно до "Кодексу академічної доброчесності ЧНУ". Політика дотримання академічної доброчесності (відповідно до Закону України "Про вищу освіту") – викладання навчальної дисципліни ґрунтується на засадах академічної доброчесності – сукупності етичних принципів та визначених законом правил, якими мають керуватися учасники освітнього процесу під час навчання, викладання та провадження наукової (творчої) діяльності з метою забезпечення довіри до результатів навчання та/або наукових (творчих) досягнень. Наявність академічного плагіату в студентських роботах є підставою для виставлення негативної оцінки. Списування студентів під час проведення модульної контрольної роботи є підставою для дострокового припинення її складання та виставлення негативної оцінки.

Складання / перескладання екзаменів відбувається за встановленим деканатом розкладом.

**Відвідування занять.** Відвідування занять є обов'язковою умовою виконання навчального плану дисципліни. Форми навчання визначені затвердженим графіком освітнього процесу Чернівецького національного університету імені Юрія Федьковича.

### **8. Рекомендована література – основна**

1. Гнатюк В. Вступ до R на прикладах: навчальний посібник.- Навчальний посібник. ХНЕУ, 2010, 107с.
2. Jenine K. Harris, Statistics With R .Washington University in St.Louis, USA, 2020.
3. Demidenko E. Advanced Statistics with Applications in R, , 2020, Wiley Series in Probability and Statistics, John Wiley and Sons USA.- 880 p.
4. Майборода Р.Є Комп'ютерна статистика. Професійний старт. Навчальний посібник. Київський університет», 2018. – 482 с.  
<http://probability.univ.kiev.ua/userfiles/mre/compsta1.pdf>
5. Королюк В.С., Царков Є.Ф., Ясинський В.К. Ймовірність, статистика та випадкові процеси. Теорія та комп'ютерна практика. В 3-х томах. Т.2: Статистика. Комп'ютерне статистичне моделювання. – Чернівці: Видавництво "Золоті литаври", 2008. – 580 с.
6. Турчин В.М. Теорія ймовірностей і математична статистика,- Дніпропетровськ, ІМА-прес, 2014. - 566 с.
7. Keon-Woong Moon. Learn ggplot2 Using Shiny App, - Springer, 2017.-P. 356.
8. Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani. An Introduction to Statistical Learning, - Springer: New , 2017.
9. Ugarte M.D., Militino A.F., Arnholt A.T. Probability and statistics with R. – Boca Raton, London, New York: CRC Press, Taylor&Francis Group, 2008. – 700 p.
10. Venables W.N., Ripley B.D. Modern applied statistics with S. – Springer, 2002. – 495 p.
11. Sarah Stowell. Using R for Statistics. Apress, 2014.

### **Додаткова**

1. Thomas Rahlf. Data Visualisation with R. Springer International Publishing, New York, 2017.

### **Методичне забезпечення**

1. Дорошенко І.В., Антонюк С.В., Бодрик Н.П. Статистика: програма, методичні вказівки і контрольні завдання. - Чернівці: Видавничий дім «Родовід» , 2013 – 76с.

### **10. Інформаційні ресурси**

1. <http://cran.r-project.org/bin/windows/base/>
2. Дистанційний курс «Аналіз даних мовою R». Круглова Н.В., Диховичний О.О.  
<https://do.ipk.kpi.ua/login/index.php>