

МІНІСТЕРСТВО ОСВІТИ І НАУКИ УКРАЇНИ  
ЛЬВІВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ ІМЕНІ ІВАНА ФРАНКА  
Кафедра статистики

**“ЗАТВЕРДЖУЮ”**  
Декан  
економічного факультету  
доц. Михайлишин Р.В.

“ \_\_\_\_\_ ” \_\_\_\_\_ 2020 року

РОБОЧА ПРОГРАМА НАВЧАЛЬНОЇ ДИСЦИПЛІНИ

**ПП 2.13 Аналіз даних в R**

\_\_\_\_\_ (шифр і назва навчальної дисципліни)  
спеціальність 051 Економіка  
\_\_\_\_\_ (шифр і назва спеціальності)  
спеціалізація “Економіка та правове регулювання в бізнесі” 051(Ю)  
\_\_\_\_\_ (назва спеціалізації)  
факультет економічний  
\_\_\_\_\_ (назва інституту, факультету, відділення)

Робоча програма дисципліни «Аналіз даних в R» для студентів  
(назва навчальної дисципліни)  
за спеціальністю 051 Економіка.

Розробники: Марець О.Р., к.е.н., доцент кафедри статистики  
(вказати авторів, їхні посади, наукові ступені та вчені звання)

Робочу програму схвалено на засіданні кафедри статистики

Протокол від «28» серпня 2020 року № 1

Завідувач кафедри статистики, професор

\_\_\_\_\_ (Матковський С.О.)  
(підпис)

«28» серпня 2020 року

© Марець О.Р., 2020 рік  
© Львівський національний  
університет, 2020

## 1. Опис навчальної дисципліни

Найменування показників	Галузь знань, освітньо-кваліфікаційний рівень	Характеристика навчальної дисципліни	
		денна форма навчання	заочна форма навчання
Кількість кредитів – 3	Галузь знань 05 Соціальні та поведінкові науки (шифр і назва)	вибіркова	
Модулів – 1	Спеціальність: 051 Економіка	Рік підготовки	
Змістових модулів – 3		2020-й	-й
Індивідуальне науково-дослідне завдання _____ (назва)		Семестр	
Загальна кількість годин – 90		IV-й	-й
Тижневих годин для денної форми навчання: аудиторних – 3 самостійної роботи студента – 2,625		Лекції	
	год.	год.	
	Практичні, семінарські		
	год.	год.	
	Лабораторні		
	48 год.	год.	
Самостійна робота			
42 год.	год.		
Індивідуальні завдання:			
год.			
Вид контролю:			
залік			

Примітка.

Співвідношення кількості годин аудиторних занять до самостійної і індивідуальної роботи становить (%):

для денної форми навчання – 114,3

## 2. Мета та завдання навчальної дисципліни

Мета: формування у студентів знань про методологію теорії статистики, а також умінь і навиків їх застосування для розв'язання дослідницьких та управлінських проблем з використанням програмного забезпечення з відкритим кодом R.

Завдання:

- розвинути знання про методи описової статистики, які доречно застосовувати для різних типів та шкал даних;
- навчитись застосовувати індуктивну статистику для дослідження вибірових сукупностей;
- оволодіти практичними навичками вирішення практичних статистичних задач в середовищі програмування R.

У результаті вивчення навчальної дисципліни студент повинен:

знати:

- основні визначення та формули описової та індуктивної статистики;
- принципи роботи в R та Rstudio та функції для вирішення практичних задач;

вміти:

- робити висновки щодо властивостей спостережуваних якісних та кількісних даних;
- визначати інтервальні оцінки невідомих параметрів генеральної сукупності та перевіряти статистичні гіпотези;
- застосувати методи багатофакторної та логістичної регресії та інтерпретувати їх результати.

## 3. Програма навчальної дисципліни

### ЗМІСТОВИЙ МОДУЛЬ 1. Основи роботи в R та RStudio

#### ТЕМА 1. Основи роботи в R та RStudio

Вікна R та Rstudio. Файл з кодом. Середовище. Консоль

Організація роботи в R та RStudio. Вікна RStudio. Створення файлу з кодом в RStudio. Основні комбінації клавіш в RStudio. Завантаження та інсталяція основних бібліотек для роботи в R.

#### ТЕМА 2. Вступ в R

Змінні. Створення та присвоєння назв змінним. Типи даних: кількісні, якісні, логічні. Вектори. Створення, присвоєння назв, вибір даних (індексування), дії з векторами. Фактори. Впорядковані та невпорядковані фактори. Дії з факторами.

Таблиці з даними (data frames). Створення, присвоєння назв, вибір даних (індексування), дії з датафреймами.

## **ЗМІСТОВИЙ МОДУЛЬ 2. Описова статистика в R**

### **ТЕМА 3. Описова статистика**

Типи даних. Показники центру розподілу та варіації. Гістограми та боксплоти. Асиметрія. Нормальний розподіл. Z-значення. Функції для описової статистики в R.

### **ТЕМА 4. Кількісні двовимірні дані**

Діаграма розсіяння. Коефіцієнт кореляції. Істотність коефіцієнта кореляції. Рівняння регресії. Регресійна таблиця. Показники якості регресійного рівняння.

## **ЗМІСТОВИЙ МОДУЛЬ 3. Статистичний висновок в R**

### **ТЕМА 5. Якісні двовимірні дані.**

Візуалізація якісних двовимірних даних. Таблиці спряженості. Умовна ймовірність. Перевірка незалежності.

### **ТЕМА 6. Лінійні функції динаміки**

Поняття функції. Метод найменших квадратів. Залишок. Коефіцієнт детермінації. R-квадрат. Поняття регресії. Застосування регресійного аналізу для вивчення трендів.

### **ТЕМА 7. Експоненційні та логістичні функції**

Експоненційна функція. Початкове значення (initial value) та коефіцієнт зростання (growth factor). Логістична функція. Вибір найкращої моделі.

### **ТЕМА 8. Центральна гранична теорема**

Розподіл вибіркової сукупності. Центральна гранична теорема. Довірчі інтервали.

### **ТЕМА 9. Гіпотези**

Умови застосування методу перевірки гіпотез для 1-ї вибірки. Типи помилок при перевірці гіпотез. Рівень альфа та критичне значення критерію.

Умови застосування методу перевірки гіпотез для 2-х вибірок. Перевірка на незалежність. Перевірка на відповідність розподілу.

### **ТЕМА 10. Хі-квадрат**

Умови застосування методу перевірки гіпотез для якісних даних. Перевірка на незалежність для якісних даних. Перевірка на відповідність розподілу для якісних даних.

### **ТЕМА 11. Дисперсійний аналіз**

Основні поняття дисперсійного аналізу та умови його застосування. F-тест. Односторонній дисперсійний аналіз. Двосторонній дисперсійний аналіз.

### **ТЕМА 12. Діагностика рівняння регресії**

Основні поняття багатофакторної регресії. Вибір чинників. Діагностика рівняння регресії. Основні поняття логістичної регресії. Класифікаційна матриця. Крива ROC.

#### 4. Структура 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>Модуль 1</b>												
<b>Змістовий модуль 1. Основи роботи в R та RStudio</b>												
Тема 1. Основи роботи в R та RStudio	7,5			4		3,5						
Тема 2. Вступ в R	7,5			4		3,5						
Разом за змістовим модулем 1	15,0			8		7,0						
<b>Змістовий модуль 2. Описова статистика в R</b>												
Тема 3. Описова статистика	7,5			4		3,5						
Тема 4. Кількісні двовимірні дані	7,5			4		3,5						
Тема 5. Якісні двовимірні дані	7,5			4		3,5						
Тема 6. Лінійні функції динаміки	7,5			4		3,5						
Тема 7. Експоненційні та логістичні функції	7,5			4		3,5						
Тема 8. Центральна гранична теорема	7,5			4		3,5						
Тема 9. Гіпотези	7,5			4		3,5						
Разом за змістовим модулем 2	52,5			28		24,5						
<b>Змістовий модуль 3. Статистичний висновок в R</b>												
Тема 10. Хі-квадрат	7,5			4		3,5						
Тема 11. Дисперсійний аналіз	7,5			4		3,5						
Тема 12. Діагностика рівняння регресії	7,5			4		3,5						
Разом за змістовим модулем 3	22,5			12		10,5						
Усього годин												
Модуль 2												
ІНДЗ			-	-		-			-	-	-	
Усього годин	90			48		42						

#### 7. Теми лабораторних занять

№ з/п	Назва теми	Кількість годин
1	Основи роботи в RStudio	4
2	Структурні частини мови програмування R	4
3	Описова статистика в R	4
4	Кількісні двовимірні дані в R	4
5	Якісні двовимірні дані в R	4
6	Лінійні функції динаміки в R	4
7	Експоненційні та логістичні функції в R	4
8	Центральна гранична теорема в R	4
9	Гіпотези в R	4

№ з/п	Назва теми	Кількість годин
10	Хі-квадрат в R	4
11	Дисперсійний аналіз в R	4
12	Діагностика рівняння регресії в R	4
	Разом	48

### 8. Самостійна робота

№ з/п	Назва теми	Кількість годин
1	Основні правила синтаксису в R.	3,5
2	Вправи swirl	3,5
3	Описова статистика в R	3,5
4	Кількісні двовимірні дані в R	3,5
5	Якісні двовимірні дані в R	3,5
6	Лінійні функції динаміки в R	3,5
7	Експоненційні та логістичні функції в R	3,5
8	Центральна гранична теорема в R	3,5
9	Гіпотези в R	3,5
10	Хі-квадрат в R	3,5
11	Дисперсійний аналіз в R	3,5
12	Діагностика рівняння регресії в R	3,5
	Разом	42

### 9. Індивідуальні завдання

#### 10. Методи навчання

Основними методами навчання, що використовуються в процесі викладання навчальної дисципліни “Аналіз даних в R” є:

- метод практичного засвоєння курсу з допомогою складання тестових завдань, вирішення задач, для набування умінь і практичних навичок (лабораторні заняття);
- метод модульного контролю з допомогою періодичного складання модулів за тематикою лекційних та практичних занять;
- метод самостійного засвоєння студентами навчального матеріалу у вигляді розв'язання тестів, вирішення задач, на підставі самостійно опрацьованої базової літератури та додаткових джерел інформації для конкретизації й поглиблення базових знань, необхідних умінь та практичних навичок (самостійна робота).

#### 11. Методи контролю

Для перевірки якості підготовки, знань, умінь студента з дисципліни використовуються такі засоби оцінювання:

- для поточного контролю - усне опитування, тестування, розв'язок прикладних задач;
- для проміжного контролю – проведення модульних завдань, що включають тестування та задачі;

- для підсумкового контролю – проведення планового заліку (тестові завдання, задачі).

## 12. Розподіл балів, які отримують студенти

Поточне тестування та самостійна робота												Сума
Змістовий модуль 1		Змістовий модуль 2							Змістовий модуль 3			
T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	T11	T12	
8,33	8,33	8,33	8,33	8,33	8,33	8,33	8,33	8,33	8,33	8,33	8,33	

T1, T2 ... T9 – теми змістових модулів.

## Шкала оцінювання: національна та ЄКТС

Оцінка ЄКТС	Сума балів за всі види навчальної діяльності	Оцінка за національною шкалою
		для заліку
A	90 – 100	зараховано
B	81-89	
C	71-80	
D	61-70	
E	51-60	
FX	21-50	не зараховано з можливістю повторного складання
F	0-20	не зараховано з обов'язковим повторним вивченням дисципліни

## 13. Методичне забезпечення

Методичне забезпечення дисципліни “Аналіз даних в R” включає:

- конспект лекцій, розміщений в системі електронного навчання Moodle;
- вказівки щодо виконання лабораторних та самостійних робіт, які розміщені в системі електронного навчання Moodle та на YouTube;
- завдань тестового типу, які супроводжують вказівки щодо **виконання** лабораторних та самостійних робіт (системі електронного навчання Moodle).
- методичних рекомендацій та завдань для виконання аудиторної та самостійної роботи на паперовому та електронному носіях.

## 14. Рекомендована література

### Базова

1. OpenIntro Statistics - Fourth Edition. David M. Diez, Harvard School of Public Health. Christopher D. Barr, Harvard School of Public Health. Link: <https://www.openintro.org/>

2. Бахрушин В.Є. Методи аналізу даних: навч. посіб. для студентів / В. Є. Бахрушин - Запоріжжя: КПУ, 2011. – 268 с.



3. Масовий онлайн навчальний курс «Аналіз даних та статистичне виведення на мові R». – Режим доступу: <https://edx.prometheus.org.ua/courses/>

4. Масовий онлайн навчальний курс «Машинне навчання». – Режим доступу: <https://edx.prometheus.org.ua/courses/>

5. R for Data Science by Hadley Wickham, Garrett Grolemund [en]. URL: <https://r4ds.had.co.nz/>

### **Допоміжна**

1. Марець О.Р. Підбір та діагностика багатофакторних регресійних моделей у пакеті аналізу даних R / О.Р.Марець, О.П.Дуда. – Режим доступу:

[http://www.ej.kherson.ua/journal/economic\\_24/2/37.pdf](http://www.ej.kherson.ua/journal/economic_24/2/37.pdf)

2. Моделювання показників зовнішньої торгівлі товарів України та Німеччини / О. Р. Марець, К. А. Михалець // Глобальні та національні проблеми економіки. Електронне фахове видання. Миколайвський національний університет імені В.О.Сухомлинського. – 2017. – Випуск 19. - Режим доступу: <http://global-national.in.ua/archive/19-2017/111.pdf>

3. R in Action, Second Edition. Data analysis and graphics with R. URL:

[https://www.manning.com/books/r-in-action-second-edition?a\\_bid=5c2b1e1d&a\\_aid=RiA2ed](https://www.manning.com/books/r-in-action-second-edition?a_bid=5c2b1e1d&a_aid=RiA2ed)

### **15. Інформаційні ресурси**

1. Sean Kross, Nick Carchedi, Bill Bauer and Gina Grdina (2020). swirl: Learn R, in R. R package version 2.4.5. <https://CRAN.R-project.org/package=swirl>

2. R packages by Hadley Wickham and Jenny Bryan. URL: <https://r-pkgs.org/>

3. Advanced R by Hadley Wickham. URL: <https://adv-r.hadley.nz/>

4. Statistical tools for high-throughput data analysis. URL: <http://www.sthda.com/>