



Аналіз соціальних даних в середовищі R

Робоча програма навчальної дисципліни (Силабус)

Реквізити навчальної дисципліни

Рівень вищої освіти	Третій (освітньо-науковий)
Галузь знань	C - соціальні науки, журналістика, інформація та міжнародні відносини
Спеціальність	C5 Соціологія
Освітня програма	Соціологія
Статус дисципліни	Вибіркова
Форма навчання	очна(денна)
Рік підготовки, семестр	2 курс, осінній семестр
Обсяг дисципліни	4 кред. ЄКТС/120 годин лекції 16 год., практичні заняття 30 год., самостійна робота 74 год.
Семестровий контроль/ контрольні заходи	Залік, МКР
Розклад занять	https://schedule.kpi.ua/
Мова викладання	Українська
Інформація про керівника курсу / викладачів	Лекції, семінарські: доктор педагогічних наук, професор Панченко Любов Феліксівна, +380963352397, lubov.felixovna@gmail.com
Розміщення курсу	https://do.ipk.kpi.ua

Програма навчальної дисципліни

1. Опис навчальної дисципліни, її мета, предмет вивчення та результати навчання

Навчальна дисципліна «Аналіз соціальних мереж» вивчається протягом одного семестру. Навчальна дисципліна належить до циклу професійної підготовки. Статус навчальної дисципліни – вибіркова. Обсяг навчальної дисципліни – 4 кредити ЄКТС.

Студенти мають отримати знання та вміння щодо мережевого аналізу, інтерпретації результатів такого аналізу. У рамках курсу вивчаються сучасні інформаційні технології та засоби візуалізації мереж.

Студенти після засвоєння дисципліни зможуть:

- орієнтуватися в історії становлення та розвитку підходів до вивчення соціальних мереж;
- використовувати базові поняття та категорії аналізу соціальних мереж;
- застосовувати основні методи та техніки дослідження соціальних мереж у соціальних науках;
- працювати з комп'ютерними інструментами для аналізу мережевих даних (Gephi, NetLogo, R, NodeXL, PSPP);
- проводити комплексні дослідження соціальних мереж різних типів (міжособистісних, організаційних, цифрових тощо);
- здійснювати обґрунтований вибір методів мережевого аналізу відповідно до дослідницьких цілей і завдань.

Засвоєння кредитного модуля сприяє підсиленню таких компетенцій та програмних результатів навчання:

- ЗК01 Здатність до абстрактного мислення, аналізу та синтезу
- ФК01 Здатність аналізувати соціальні явища і процеси
- ФК12 Здатність застосовувати сучасні статистичні методи, моделі, цифрові технології, спеціалізоване програмне забезпечення для моделювання соціальних процесів
- ПР01 Аналізувати соціальні явища і процеси, використовуючи емпіричні дані та сучасні концепції і теорії соціології
- ПР13 Застосовувати сучасні новітні методи соціологічних досліджень в умовах цифровізації суспільних відносин

2. Пререквізити та постреквізити дисципліни (місце в структурно-логічній схемі навчання за відповідною освітньою програмою)

Навчальна дисципліна спирається на знання студентів з курсу «Актуальні питання методології соціологічних досліджень»,

3. Зміст навчальної дисципліни

1. R як середовище для аналізу даних
2. Структури даних та імпорт
3. Філософія tidyverse і пакет dplyr
4. Очищення й трансформація даних (tidyr). Граматика графіків.
5. Описова статистика та поглиблена візуалізація в R
6. Вступ до статистичного висновку
7. Інтерпретація та припущення лінійної регресії
8. Багатовимірна статистика в R. Відтворювані звіти

4. Навчальні матеріали та ресурси

Рекомендована література

Базова

1. Кофанов, О. Є., Солнцев, С. О., & Зозульов, О. В. (2023). Програмування із використанням R у статистичних та маркетингових дослідженнях: Навчально-методичний комплекс дисципліни. КПІ ім. Ігоря Сікорського.
2. Field, A., Miles, J., & Field, Z. (2012). *Discovering statistics using R*. SAGE Publications.
3. Kabacoff, R. I. (2022). *R in action: Data analysis and graphics with R and Tidyverse* (3rd ed.). Manning Publications.
4. Navarro, D. (2019). *Learning statistics with R: A tutorial for psychology students and other beginners*. <https://learningstatisticswithr.com>
5. Wickham, H., Çetinkaya-Rundel, M., & Grolemund, G. (2023). *R for data science: Import, tidy, transform, visualize, and model data* (2nd ed.). O'Reilly Media.

Допоміжна

1. Панченко Л.Ф. (2018) До питання використання кількісних методів у дослідженнях конфліктів. *Вісник Національного технічного університету України "Київський політехнічний інститут". Політологія. Соціологія. Право* : зб. наук. праць. Київ, No 3 (39). С.21–27.
2. Панченко Л.Ф. (2018). *Математичні та статистичні методи аналізу соціологічної інформації*. Київ, КПІ ім.Ігоря Сікорського.
3. Fox, J., & Weisberg, S. (2019). *An R companion to applied regression* (3rd ed.). SAGE Publications.

4. Llaudet, E., & Imai, K. (2022). *Data analysis for social science: A friendly and practical introduction*. Princeton University Press.
5. Imai, K. (2017). *Quantitative social science: An introduction*. Princeton University Press.

Навчальний контент

5. Методика опанування навчальної дисципліни (освітнього компонента)

Лекційні заняття

Лекція 1. Знайомство з R як середовищем для аналізу даних

R це вільна мова та середовище для статистичних обчислень і графіки, що виросло з мови S і сьогодні спирається на екосистему CRAN (понад 20 тисяч пакетів). Її сильні сторони для соціолога: безкоштовність і кросплатформність, відтворюваність аналізу (скрипт замість «кліків»), потужна графіка та велика спільнота; слабкі сторони: крутіша, ніж у SPSS, початкова крива навчання, зберігання даних переважно в оперативній пам'яті та потреба керувати версіями пакетів. На занятті розглядаються чотири панелі RStudio (консоль, редактор скриптів, оточення та історія, файли, графіки, пакети, довідка), базові типи об'єктів (числові, символьні, логічні), арифметичні оператори з урахуванням пріоритету, виклик функцій з іменованими аргументами та способи отримання довідки. Ключовий дидактичний акцент: із першого заняття привчати писати відтворюваний скрипт з коментарями та працювати в межах RStudio-проєкту, а не лише в консолі; це закладає культуру відтворюваності на весь курс.

Лекція 2. Структури даних та імпорт

Лекція формує «граматику» даних R: атомарні вектори (однорідні), фактори (категоріальні змінні з рівнями та мітками, що є критичним для порядкових шкал анкет), матриці, датафрейми (основна прямокутна структура для даних опитувань) і списки (неоднорідні, вкладені; саме так влаштовані результати статистичних моделей). Детально відпрацьовується індексация (за позицією, логічна, за іменами) і правила приведення типів. Друга частина присвячена імпорту: `read.csv` проти `readr::read_csv`, Excel через `readxl`, файли SPSS (.sav) через `haven` зі збереженням міток значень і змінних, що особливо важливо для масивів ESS, WVS, ISSP чи PIAAC. Для тих, хто переходить із SPSS, окремо показано пакети `labelled` і `sjlabelled`, які дають змогу зручно працювати з мітками у звичному стилі та перетворювати їх на фактори. Обговорюються також кодування (UTF-8, кирилиця) і типові помилки імпорту.

Лекція 3. Філософія tidyverse і пакет dplyr

Лекція вводить концепцію «охайних даних» (tidy data): кожна змінна це стовпець, кожне спостереження це рядок, кожне значення це комірка, і показує tidyverse як узгоджений набір пакетів, побудований навколо цього принципу. Центральне місце посідають дієслова dplyr: `select()` (відбір стовпців), `filter()` (відбір рядків), `mutate()` (створення та перекодування змінних), `arrange()` (впорядкування), а також оператор конвеєра `%>%` (та нативний `|>`), що робить код читабельним зліва направо на противагу вкладеним викликам базового R. На прикладах даних опитувань демонструється, як ця логіка відповідає типовим завданням підготовки соціологічних даних, і закладається підґрунтя для згрупованих обчислень наступних лекцій.

Лекція 4. Очищення й трансформація даних (tidyr) та вступ до ggplot2

Перша частина присвячена формі даних: відмінності «довгого» й «широкого» форматів та функціям `pivot_longer()` і `pivot_wider()` (сучасні наступники `gather()` і `spread()`), потрібним для панельних і повторюваних вимірів та для побудови графіків за багатьма змінними. Розглядаються пропущені значення (NA): їх типи, виявлення (`is.na()`, `complete.cases()`), аргумент `na.rm` і базові стратегії обробки; окремо зазначено, що вилучення за списком є найпростішим, але не єдиним

шляхом, і за потреби застосовують множинну імпутацію (пакет `mise`). Друга частина знайомить із граматикою графіки `ggplot2`: дані плюс естетики (`aes`) плюс геомі плюс фасети плюс шкали плюс теми як шарова конструкція. Здобувачі будують перші графіки й засвоюють, що `ggplot2` це мова опису візуалізації, а не набір готових «типів діаграм».

Лекція 5. Описова статистика та поглиблена візуалізація

Лекція систематизує описовий аналіз: частоти (`table()`, `count()`), таблиці спряженості з частками й підсумками, показники центральної тенденції (середнє, медіана, мода) та розсіювання (стандартне відхилення, дисперсія, IQR, квартилі) з наголосом на відповідності показника рівню вимірювання змінної. Ключова соціологічна навичка це агрегація за групами через `group_by()` і `summarise()` для порівняння підвбірок (за статтю, віком, освітою, регіоном). Оскільки соціологічні дослідження зазвичай спираються на складні вибірки, окремо вводиться зважене оцінювання: пакети `survey` і `srvyr` для коректних середніх і часток та `questionr` для зважених таблиць спряженості; ігнорування ваг у базовому аналізі спотворює репрезентативні оцінки. Друга частина розвиває `ggplot2` до якості, придатної для публікації: заголовки, підписи осей і приміток, керування кольором і заливкою, теми, фасетування за підгрупами; згадуються готові рішення пакета `sjPlot` для соціальних даних.

Лекція 6. Вступ до статистичного висновку

Лекція вибудовує логіку висновку: генеральна сукупність і вибірка, вибіркові розподіли, нульова гіпотеза, р-значення, рівень значущості, довірчі інтервали та розмір ефекту, із сучасним акцентом на тому, що інтервали й розміри ефекту інформативніші за «голе» р-значення (пакет `effectsize` для стандартизованих мір). Розглядаються кореляція (Пірсона й Спірмена, умови застосування), t-тест (одновибірковий, для незалежних і залежних вибірок), критерій хі-квадрат для зв'язку категоріальних змінних та перша лінійна регресія через `lm()`. Наголошується на обережності з причиновою інтерпретацією кореляційних і регресійних результатів на неекспериментальних соціологічних даних.

Лекція 7. Інтерпретація та припущення лінійної регресії

Заняття поглиблює читання результатів `lm()`: нестандартизовані й стандартизовані (бета) коефіцієнти, стандартні похибки, t-значення та р-значення, R-квадрат і скоригований R-квадрат, F-критерій моделі, і, головне, змістовна інтерпретація коефіцієнтів у соціологічних термінах. Детально розбираються припущення регресії (лінійність, незалежність спостережень, гомоскедастичність, нормальність залишків, відсутність мультиколінеарності) та їх діагностика графіками (`plot(model)`: залишки проти передбачених, Q-Q, масштаб-локація, важелі й відстань Кука) і показником VIF. Розглядається кодування категоріальних предикторів (`dummy`) і коротко ефекти взаємодії. Оскільки бінарні залежні змінні в соціології поширені, вводиться логістична регресія (`glm`, родина `binomial`) як природне розширення; для оформлення регресійних таблиць «під публікацію» показано пакети `sjPlot` і `gtsummary`.

Лекція 8. Реалізація багатовимірних методів у R та відтворювані звіти для соціолога

Завершальна лекція має дві частини. Перша: багатовимірні методи (дисперсійний, дискримінантний, кластерний і факторний аналіз), теоретично опановані здобувачами в окремому курсі, розглядаються тут з боку реалізації в R і закріплюються на практичних заняттях 11–13 (функції `aov`, `kmeans`, `hclust` та пакети `MASS` і `psych`). Друга частина: відтворюване звітування засобами RMarkdown, коли наратив, код і результати поєднано в одному документі з виводом у HTML або Word. Окремо подано екосистему пакетів, зручних саме для соціолога: `survey` і `srvyr` для зважених оцінок за складними вибірками, `questionr` для зважених таблиць спряженості, `labelled` і `sjlabelled` для роботи з мітками у стилі SPSS, `sjPlot` і `sjmisc` для таблиць та графіків соціальних даних, `gtsummary` для публікаційних таблиць, `psych` для шкал і оцінювання надійності. Пакет `lavaan` для

структурного моделювання згадується як місток до профільного курсу зі структурного моделювання.

Практичні заняття

Практичне заняття 1. Знайомство з R і RStudio

Практичне заняття 2. Структури даних та індексація

Питання, що розглядаються на занятті:

1. Вектори, фактори, матриці, датафрейми, списки: призначення кожної структури.
2. Фактори для номінальних і порядкових змінних анкети (рівні, мітки, впорядкованість).
3. Індексація: за позицією, логічна, за іменами.
4. Приведення типів (coercion) у змішаних даних.

Завдання для самостійної роботи:

1. Створити вручну датафрейм із кількох змінних (стать, вік, освіта, дохід) на 8–10 спостережень.
2. Перетворити порядкову змінну освіти на впорядкований фактор із коректними рівнями.
3. Витягти підмножину респондентів за логічною умовою (наприклад, жінки з вищою освітою).
4. Пояснити письмово, який тип має кожен стовпець і чому.

Практичне заняття 3. Імпорт соціологічних даних та їх первинний огляд

Питання, що розглядаються на занятті:

1. Читання CSV, Excel і SPSS (.sav): `read_csv` (readr), `read_excel` (readxl), `read_sav` (haven).
2. Збереження міток змінних і значень; робота з пакетами `labelled` і `sjlabelled` для даних у стилі SPSS.
3. Первинний огляд даних: `str()`, `summary()`, `head()`, `tail()`.
4. Кодування (UTF-8, кирилиця) та типові помилки імпорту.

Завдання для самостійної роботи:

1. Імпортувати наданий набір даних опитування у форматах .csv та .sav і порівняти результати.
2. Оглянути структуру даних (`str()`, `summary()`), визначити типи всіх змінних.
3. Перетворити щонайменше одну мічену змінну (`labelled`) на фактор і перевірити результат.
4. Скласти короткий «паспорт даних»: кількість спостережень, змінних, їх типи, наявність пропусків.

Практичне заняття 4. Основи dplyr: select, filter, mutate, arrange, конвеєр

Питання, що розглядаються на занятті:

1. Дієслова `select()`, `filter()`, `mutate()`, `arrange()` та їх призначення.
2. Оператор конвеєра `%>%` (`|>`) і читабельність коду.
3. Створення та перекодування змінних через `mutate()`.
4. Порівняння підходу `dplyr` з базовим R.

Завдання для самостійної роботи:

1. На даних опитування відібрати релевантні змінні та відфільтрувати підгрупу (наприклад, 25–45 років).
2. Створити перекодовану змінну (вікові групи) через `mutate()`.
3. Упорядкувати результат за доходом за спаданням.
4. Записати весь аналіз одним конвеєром і прокоментувати кроки.

Практичне заняття 5. Конвеєри dplyr на реальних соціальних даних

Питання, що розглядаються на занятті:

1. Побудова багатокрокових конвеєрів для відповіді на змістовне питання.
2. Перекодування через `case_when()`; категоризація неперервних змінних.
3. Перевірка коректності перекодування таблицями частот.
4. Швидкі (у тому числі зважені) таблиці спряженості засобами `questionr`.

Завдання для самостійної роботи:

1. Сформулювати змістовне питання (наприклад, середній дохід за рівнем освіти серед зайнятих).
2. Побудувати багатокроковий конвеєр, що відповідає на це питання.
3. Перекодувати щонайменше одну змінну через `case_when()` і перевірити таблицею частот.
4. Стисло описати отриманий результат у соціологічних термінах.

Практичне заняття 6. Об'єднання таблиць і групування даних

Питання, що розглядаються на занятті:

1. Типи з'єднань (`left_join`, `inner_join`, `full_join`) та роль ключової змінної.
2. Групування `group_by()` і зведення `summarise()`.
3. Підрахунок частот `count()`, у тому числі за кількома ознаками.
4. Підготовка агрегованих таблиць для порівняння підгруп.

Завдання для самостійної роботи:

1. Об'єднати таблицю респондентів із таблицею регіональних показників за ключем регіону.
2. Обчислити середні значення обраної змінної за групами (регіон, стать) через `group_by()` і `summarise()`.
3. Побудувати таблицю частот за двома ознаками за допомогою `count()`.
4. Пояснити, які змістовні відмінності між групами виявлено.

Практичне заняття 7. Трансформація форми даних (tidyr) і базова візуалізація (ggplot2)

Питання, що розглядаються на занятті:

1. «Довгий» і «широкий» формати даних; `pivot_longer()` і `pivot_wider()`.
2. Обробка пропущених значень; стратегії та коротко про множинну імпутацію (`mice`).
3. Базові геомі: гістограма, точковий графік, стовпчикова діаграма, боксплот.
4. Логіка граматики графіки `ggplot2`.

Завдання для самостійної роботи:

1. Перетворити набір змінних у «довгий» формат для подальшої візуалізації.
2. Опрацювати пропущені значення обраною стратегією та обґрунтувати вибір.
3. Побудувати чотири базові графіки (`geom_histogram`, `geom_point`, `geom_bar`, `geom_boxplot`).
4. Стисло проінтерпретувати кожен графік.

Практичне заняття 8. Статистики за групами та поглиблена візуалізація

Питання, що розглядаються на занятті:

1. Кілька зведених статистик одночасно в `summarise()`.
2. Зважені оцінки для репрезентативних даних: пакети `survey` і `srvyr`.
3. Фасетування та оформлення графіків «під публікацію»; можливості `sjPlot`.
4. Збереження графіків (`ggsave`) з належною роздільною здатністю.

Завдання для самостійної роботи:

1. Обчислити середнє, медіану та стандартне відхилення обраної змінної за групами.
2. Обчислити зважену оцінку середнього або частки з урахуванням ваги вибірки (survey або srvyr).
3. Побудувати фасетований графік розподілу змінної за категоріями та оформити його «під публікацію».
4. Зберегти графік у файл (ggsave) із належною роздільною здатністю.

Практичне заняття 9. Кореляція, t-тест, хі-квадрат і перша регресія

Питання, що розглядаються на занятті:

1. Кореляція Пірсона і Спірмена; умови застосування кожної.
2. t-тест (незалежні та залежні вибірки) і критерій хі-квадрат.
3. Проста лінійна регресія `lm()` та читання `summary()`.
4. Розмір ефекту й довірчі інтервали; пакет `effectsize`.

Завдання для самостійної роботи:

1. Обчислити коефіцієнти кореляції між двома неперервними змінними та проінтерпретувати їх.
2. Провести t-тест для порівняння середніх у двох групах і хі-квадрат для двох категоріальних змінних.
3. Побудувати просту лінійну регресію та описати коефіцієнти й р-значення.
4. Доповнити висновок розміром ефекту та довірчим інтервалом, а не лише значущістю.

Практичне заняття 10. Множинна лінійна та логістична регресія

Питання, що розглядаються на занятті:

1. Множинна регресія: інтерпретація коефіцієнтів і скоригованого R-квадрата.
2. Кодування категоріальних предикторів (`dummy`) та стандартизовані коефіцієнти.
3. Діагностика припущень (`plot(model)`) і мультиколінеарність (VIF).
4. Логістична регресія (`glm, binomial`) для бінарних залежних змінних; таблиці моделей засобами `sjPlot` і `gtsummary`.

Завдання для самостійної роботи:

1. Побудувати множинну регресію для соціального показника з 3–4 предикторами та проінтерпретувати коефіцієнти.
2. Перевірити припущення діагностичними графіками та показником VIF.
3. Побудувати логістичну регресію для бінарної залежної змінної та пояснити відношення шансів.
4. Оформити результати публікаційною таблицею (`sjPlot` або `gtsummary`) і прокоментувати.

Практичне заняття 11. Дисперсійний аналіз (ANOVA)

Питання, що розглядаються на занятті:

1. Однофакторна та двофакторна ANOVA; функція `aov()`.
2. F-критерій та його інтерпретація.
3. Пост-хок порівняння (тест Тьюкі).
4. Припущення (тест Левена) і розмір ефекту (η^2 -квадрат).

Завдання для самостійної роботи:

1. Порівняти середнє значення показника між трьома і більше групами через `aov()`.
2. Провести пост-хок порівняння та визначити, які групи відрізняються.
3. Перевірити припущення ANOVA й оцінити розмір ефекту (η^2 -квадрат).

4. Сформулювати змістовний висновок за результатами аналізу.

Практичне заняття 12. Дискримінантний аналіз

Питання, що розглядаються на занятті:

1. Призначення лінійного дискримінантного аналізу (LDA).
2. Відмінність LDA від логістичної регресії за призначенням.
3. Побудова моделі засобами MASS::lda та її припущення.
4. Оцінювання якості класифікації (матриця плутанини, точність).

Завдання для самостійної роботи:

1. Побудувати LDA для класифікації респондентів за груповою ознакою.
2. Оцінити точність класифікації на основі матриці плутанини.
3. Проінтерпретувати дискримінантні функції та внесок змінних.
4. Обговорити обмеження методу для даних дослідження.

Практичне заняття 13. Кластерний і факторний аналіз

Питання, що розглядаються на занятті:

1. Ієрархічна кластеризація і метод k-середніх; вибір кількості кластерів.
2. Розвідувальний факторний аналіз (psych::fa) та факторні навантаження.
3. Придатність даних для факторного аналізу: показник КМО і критерій Бартлетта.
4. Відмінність EFA від методу головних компонент (PCA).

Завдання для самостійної роботи:

1. Провести кластеризацію респондентів (k-means та або hclust) і охарактеризувати кластери.
2. Виконати розвідувальний факторний аналіз набору установочних тверджень.
3. Проінтерпретувати виділені фактори та оцінити придатність даних (КМО, Бартлетт).
4. Стисло описати змістовну структуру, виявлену аналізом.

Практичне заняття 14. Відтворювані звіти (RMarkdown) та інструментарій соціолога

Питання, що розглядаються на занятті:

1. Структура документа RMarkdown (YAML, чанки коду, вбудований код).
2. Складання (knit) звіту у форматі HTML і Word.
3. Пакети соціолога у звіті: sjPlot, gtsummary, survey, questionr, labelled.
4. Відтворюваність як принцип відкритої науки.

Завдання для самостійної роботи:

1. Створити документ RMarkdown із завантаженням даних, аналізом і графіками.
2. Додати вбудований код (inline) для автоматичного підставлення чисел у текст.
3. Оформити щонайменше одну таблицю засобами gtsummary або sjPlot.
4. Зібрати звіт у HTML і Word та переконатися, що він відтворюється «з нуля».

Практичне заняття 15. Кейс з аналізу даних (модульна контрольна робота)

Питання, що розглядаються на занятті:

1. Повний цикл аналізу даних: від дослідницького питання до звіту.
2. Обґрунтування вибору статистичного методу під питання й тип даних.
3. Коректне подання результатів з розмірами ефекту та (за потреби) вагами вибірки.

4. Принципи відтворюваності в оформленні кейсу.

Завдання для самостійної роботи:

1. На новому наборі даних сформулювати дослідницьке питання і гіпотезу.
2. Виконати наскрізний аналіз: імпорт, підготовка, описова статистика, доречний метод висновку.
3. Оформити результат як відтворюваний звіт (RMarkdown) із коментарями та інтерпретацією.
4. Подати стислий висновок із критичним обговоренням обмежень аналізу.

6. Самостійна робота студента

Самостійна робота здобувача включає:

підготовка до аудиторних занять – 64 год;

підготовка до модульної контрольної роботи – 4 год;

підготовка до заліку – 6 год.

Загалом – 74 год.

Види самостійної роботи: підготовка до практичних занять (перегляд матеріалів поточної лекції), виконання завдань для самостійної роботи. Усі терміни виконання (дедлайни) наводяться в системі Мудл потижнево.

Політика дисципліни

7. Політика навчальної дисципліни (освітнього компонента)

Навчальні заняття з дисципліни “Аналіз соціальних даних в середовищі R” проводяться у формі лекційних занять та занять комп’ютерного практикуму. На лекціях викладач знайомить студентів з ключовими поняттями, основними методами, які використовуються для аналізу даних. Лекції відбуваються в діалоговій формі з використанням мультимедійних презентацій, запитаннями та дискусією. На практичних заняттях студенти виконують завдання комп’ютерного практикуму з аналізу соціальних даних в середовищі R. Модульна контрольна робота призначена для оцінювання знань та умінь студентів щодо використання R для аналізу даних обраної предметної області, та їх візуалізації.

В ході вивчення курсу застосовуються стратегії активного і колективного навчання, які реалізуються наступними методами і технологіями: методи проблемного навчання; особистісно-орієнтовані технології, зокрема кейс-стаді з аналізу даних, інформаційно-комунікаційні технології, які активізують роботу студентів (мультимедійні презентації для лекційних занять, карти знань, інфографіка та сторітелінг).

Всі заняття супроводжуються матеріалами та завданнями в системі дистанційного навчання Moodle.

На лекціях та заняттях вітаються питання, дискусії, діалог; передбачається атмосфера толерантності та поваги до Іншого. Викладачі та студенти керуються нормами етичної поведінки та принципами академічної доброчесності у викладанні та навчання (Кодекс честі КПІ ім. Ігоря Сікорського). Студенти можуть звернутися до викладача по необхідну допомогу або консультацію face-to-face або засобами електронної пошти та соціальних мереж. Заохочувальні бали додаються до семестрового рейтингу студента, який максимально може досягати 100 балів.

Відвідування занять

Відвідування лекційних та практичних занять є обов'язковим.

Пропущені контрольні заходи оцінювання

Пропущену модульну контрольну роботу можна написати повторно, але до моменту проставлення проміжної атестації з урахуванням часу, необхідного викладачу для перевірки роботи.

Процедура оскарження результатів контрольних заходів оцінювання

Студенти мають можливість підняти будь-яке питання, яке стосується процедури контрольних заходів та очікувати, що воно буде розглянуто згідно із наперед визначеними процедурами. Для оскарження контрольних заходів студент має подати заяву, у якій має бути вказана причина оскарження, наведені факти упередженості викладача. Викладач повинен обговорити дану заяву зі студентом особисто на консультації. В разі відсутності порозуміння щодо результату контрольного заходу, формується комісія з викладачів кафедри, яка оцінює процедуру проведення контрольного заходу та претензії студента. Комісія може вирішити провести контрольний захід повторно, або відхилити заяву. Рішення комісії є остаточним та оскарженню не підлягає.

Політика університету

Академічна доброчесність

Політика та принципи академічної доброчесності визначені у розділі 3 Кодексу честі Національного технічного університету України «Київський політехнічний інститут імені Ігоря Сікорського». Детальніше: <https://kpi.ua/code>.

Норми етичної поведінки

Норми етичної поведінки студентів і працівників визначені у розділі 2 Кодексу честі Національного технічного університету України «Київський політехнічний інститут імені Ігоря Сікорського». Детальніше: <https://kpi.ua/code>.

Політика використання штучного інтелекту

Політика використання штучного інтелекту та її принципи регламентуються наказом «Політика використання штучного інтелекту для академічної діяльності в КПІ ім. Ігоря Сікорського». Детальніше: <https://osvita.kpi.ua/node/1225>.

8. Види контролю та рейтингова система оцінювання результатів навчання (PCO)

Оцінювання та контрольні заходи

Рейтинг студента з дисципліни складається з балів, що отримуються за:

1. виконання практичних робіт в комп'ютерному класі;
2. виконання МКР.

1. Виконання практичних робіт в комп'ютерному класі оцінюється у 5 балів.

Максимальна кількість балів за всі практичні роботи дорівнює $r_{\text{пр}} = 5 \text{ балів} \cdot 14 = 70 \text{ балів}$.

Критерії оцінювання:

- «відмінно», завдання виконано повністю (не менше 90% потрібної роботи), без суттєвих помилок. Студент демонструє впевнене володіння програмним забезпеченням, логічно і послідовно виконує завдання, аргументовано пояснює свої дії та відповідає на всі поставлені питання – 5 балів;
- «добре», завдання виконано достатньо повно (не менше 75% потрібної роботи) або повністю, але з незначними помилками. Студент здебільшого розуміє алгоритм роботи, допускає окремі неточності, виправляє їх після підказок викладача, відповідає на більшість поставлених питань – 3-4 бали;
- «задовільно» – завдання виконано частково (не менше 60% потрібної роботи), містить значні помилки. Студент погано орієнтується в роботі, не може самостійно виправити помилки, відповідає на питання невпевнено або не може пояснити свої дії – 1-2 бали;

- «незадовільно» – робота не виконана - 0 балів.

2. Модульна контрольна робота (кейс з аналізу даних)

Ваговий бал – 30.

Відповіді оцінюються наступним чином:

- «відмінно» – повна відповідь (не менше 90% потрібної інформації) – 28-30 балів;
- «добре» – достатньо повна відповідь (не менше 75% потрібної інформації), або повна відповідь з незначними неточностями – 24-27 балів;
- «задовільно» – неповна відповідь (не менше 60% потрібної інформації) та незначні помилки – 18-22 бали;
- «незадовільно» – відповідь не відповідає вимогам на «задовільно» – 0-6 балів.

3. Заохочувальні бали

Всього не більше 10 балів за такі види робіт:

- за науково-дослідницьку діяльність (участь у конференціях, конкурсах студентських робіт, публікації);
- участь у факультетських олімпіадах з дисципліни та всеукраїнських олімпіадах.

Рейтингова оцінка з навчальної дисципліни доводиться до здобувачів на заліку під час заліково-екзаменаційної сесії.

Здобувачі, які мають рейтингову оцінку 60 і більше балів, отримують відповідну до набраного рейтингу оцінку без додаткових випробувань.

Зі здобувачами, які мають рейтингову оцінку менше 60 балів, а також з тими здобувачами, хто бажає підвищити свою рейтингову оцінку, викладач проводить семестровий контроль у вигляді залікової контрольної роботи або співбесіди.

4. Залік.

Ваговий бал – 100.

Залік має вигляд переліку питань, на які повинен відповісти студент. Питання є різними за змістом, відповідають тематиці лекцій, семінарів, самостійної роботи, питань з самоконтролю.

Критерії оцінювання

90–100 балів – студент демонструє глибоке знання змісту навчального матеріалу, здатність аналізувати питання у міждисциплінарному контексті, вільно оперує науковими поняттями та термінами, робить обґрунтовані висновки, виявляє власну позицію з дискусійних питань.

75–89 балів – студент демонструє загальне розуміння основних тем і питань курсу, проте допускає окремі неточності у визначеннях, прикладах або висновках. Використання наукової термінології є частковим або непослідовним, висновки не завжди аргументовані.

60–74 балів – студент демонструє поверхові або фрагментарні знання, не орієнтується у ключових поняттях курсу, не проявляє аналітичного мислення чи вміння застосовувати отримані знання; відповіді нелогічні або неповні.

Умови позитивної проміжної атестації:

Для отримання «зараховано» з першої проміжної атестації студент повинен мати не менше ніж 10 балів, для отримання «зараховано» з другої проміжної атестації студент повинен мати не менше ніж 20 балів.

Умови допуску до заліку:

Умовою допуску студента до заліку є поточний контроль не менше 36 балів.

Таблиця відповідності рейтингових балів оцінкам за університетською шкалою:

Рейтингова оцінка здобувача (бали)	Університетська шкала оцінок рівня здобутих компетентностей (результатів навчання)
100-95	Відмінно
94-85	Дуже добре
84-75	Добре
74-65	Задовільно
64-60	Достатньо
Менше 60	Незадовільно

Можливі відмітки у відомості семестрового контролю:

Не допущено	Невиконання умов допуску до семестрового контролю
Усунено	Порушення принципів академічної доброчесності або морально-етичних норм поведінки
Не з'явився	Здобувач, був допущений, але не з'явився на залік

Перегляд рейтингової системи оцінювання впродовж семестру

Перегляд РСО може здійснюватися за вмотивованою заявою здобувача, що вивчає ОК, органу студентського самоврядування або профспілкового комітету студентів, поданою на ім'я завідувача забезпечуючої кафедри. Процедура перегляду визначена у розділі 7 Положення про систему оцінювання результатів навчання в КПІ ім. Ігоря Сікорського https://osvita.kpi.ua/sites/default/files/downloads/Pologennia_RSO_2025.pdf

9. Додаткова інформація з дисципліни (освітнього компонента)

Додаток 1. Перелік питань до заліку з дисципліни «Аналіз соціальних даних в середовищі R»

1. R як середовище для аналізу соціальних даних: переваги та недоліки порівняно зі спеціалізованими статистичними пакетами (SPSS). Структура інтерфейсу RStudio.
2. Типи об'єктів і базові операції в R. Функції та способи отримання довідки. Поняття відтвореного скрипту й RStudio-проєкту.
3. Основні структури даних у R (вектори, фактори, матриці, датафрейми, списки): призначення та відмінності. Роль факторів для категоріальних і порядкових змінних анкети.
4. Індексція об'єктів у R (за позицією, логічна, за іменами). Приведення типів даних.
5. Імпорт даних із форматів CSV, Excel, SPSS (.sav). Роль пакета haven; збереження міток значень і змінних; пакети labelled і sjlabelled.
6. Концепція «охайних даних» (tidy data). Пакет dplyr: основні дієслова (select, filter, mutate, arrange). Оператор конвеєра.
7. Перетворення форми даних: «довгий» і «широкий» формати, функції pivot_longer і pivot_wider; коли застосовують кожну.

8. Пропущені значення в R: типи, виявлення, стратегії обробки; поняття множинної імпутації (пакет `mice`).
9. Граматика графіки `ggplot2`: основні складники (дані, естетики, геомі, фасети, шкали, теми).
10. Описова статистика в R: показники центральної тенденції та розсіювання; відповідність показника рівню вимірювання змінної.
11. Таблиці частот і крос-табуляції в R. Агрегація даних за групами (`group_by`, `summarise`).
12. Зважений аналіз соціологічних даних: призначення ваг вибірки; пакети `survey`, `srvyr`, `questionr`.
13. Логіка статистичного висновку: генеральна сукупність і вибірка, нульова гіпотеза, р-значення, рівень значущості, довірчий інтервал.
14. Поняття розміру ефекту та його значення. Чому недостатньо орієнтуватися лише на р-значення.
15. Кореляційний аналіз у R: коефіцієнти Пірсона і Спірмена; умови застосування кожного.
16. t-тест (одновибірковий, для незалежних і залежних вибірок): призначення та припущення.
17. Критерій хі-квадрат для аналізу зв'язку між категоріальними змінними.
18. Проста лінійна регресія в R (функція `lm`): побудова моделі та інтерпретація результатів.
19. Множинна лінійна регресія: інтерпретація коефіцієнтів, R-квадрат і скоригований R-квадрат, стандартизовані (бета) коефіцієнти.
20. Припущення лінійної регресії та їх діагностика (графіки залишків, показник VIF).
21. Кодування категоріальних предикторів (`dummy`) у регресії. Поняття ефектів взаємодії.
22. Логістична регресія (`glm`, родина `binomial`): призначення, відношення шансів, відмінність від лінійної регресії.
23. Дисперсійний аналіз (ANOVA) в R: однофакторний та двофакторний, F-критерій, пост-хок порівняння, припущення.
24. Дискримінантний аналіз (LDA): призначення, припущення, оцінювання якості класифікації; відмінність від логістичної регресії.
25. Кластерний аналіз у R: ієрархічний метод і метод k-середніх; підходи до вибору кількості кластерів.
26. Розвідувальний факторний аналіз (EFA): факторні навантаження, придатність даних (КМО, критерій Бартлетта); відмінність від методу головних компонент.
27. Відтворювані звіти в R: структура документа RMarkdown, чанки коду, вивід у HTML і Word. Співвідношення RMarkdown і Quarto.
28. Екосистема пакетів для соціолога в R (`survey`, `sjPlot`, `gtsummary`, `questionr`, `labelled`, `psych`): призначення кожного.
29. Принципи відтворюваності дослідження та відкритої науки: роль скриптів, проєктів і фіксації версій пакетів (`renv`).
30. Повний цикл аналізу соціальних даних у R: від імпорту й підготовки до статистичного висновку та звіту. Обґрунтування вибору методу під дослідницьке питання й тип даних.

Додаток 2. Неформальна освіта

Існує можливість зарахування сертифікатів проходження дистанційних та онлайн курсів за тематикою дисципліни, або окремих модулів курсів. Зокрема, рекомендується навчання на освітніх платформах:

1) Data Camp: курси Introduction in R, Introduction to statistics with R, Statistics fundamental with R, case study Network Analysis in R

(<https://campus.datacamp.com/courses/case-studies-network-analysis-in-r/exploring-graphs-through-time?ex=1>

<https://campus.datacamp.com/courses/case-studies-network-analysis-in-r/exploring-graphs-through-time?ex=6>)

2) Coursera: Specialization Computational Social Science

(<https://www.coursera.org/specializations/computational-social-science-ucdavis>) та курс

Network Dynamics of Social Behavior (<https://www.coursera.org/learn/networkdynamics>)

Дистанційне навчання

Можливе синхронне дистанційне навчання з використанням платформ для відео-конференцій та освітньої платформи для дистанційного навчання в університеті.

Інклюзивне навчання

Допускається

Робочу програму навчальної дисципліни (силабус):

Складено професором кафедри, доктором педагогічних наук, професором Панченко Л.Ф.

Ухвалено кафедрою соціології (протокол № 16 від 23.06.2026)

Погоджено Методичною комісією факультету соціології і права (протокол № 7 від 30.06.2026)