

Волинський національний університет імені Лесі Українки
Факультет інформаційних технологій і математики
Кафедра загальної математики та методики навчання інформатики

Марія Хомяк

МАТЕМАТИЧНІ МЕТОДИ ОБРОБКИ ДАНИХ
ПСИХОЛОГО-ПЕДАГОГІЧНИХ ДОСЛІДЖЕНЬ

Лабораторний практикум для здобувачів освіти спеціальності

014 Середня освіта (Інформатика)

другого (магістерського) рівня освіти

Луцьк 2025

УДК 004:51-7-047.37(076)

X 76

Рекомендовано до друку науково-методичною радою Волинського національного університету імені Лесі Українки (протокол №7 від 19.03.2025р.)

Рецензент:

С.М. Яцюк, к.пед. н., доц., декан факультету інформаційних технологій і математики Волинського національного університету імені Лесі Українки

Хомяк М.Я.

X 76 Математичні методи обробки даних психолого-педагогічних досліджень: лабораторний практикум для здобувачів освіти спеціальності 014 Середня освіта (Інформатика) другого (магістерського) рівня освіти [Електронний ресурс] / Марія Ярославівна Хомяк. Луцьк : ВНУ ім. Лесі Українки, 2025. 32 с.

Анотація: Лабораторний практикум містить ряд лабораторних робіт з освітнього компонента «Математичні методи обробки даних психолого-педагогічних досліджень». Матеріал структуровано за темами, до кожної із тем наведено відповідні завдання. Видання призначене для здобувачів освіти, що навчаються за спеціальністю 014 Середня освіта (Інформатика) другого (магістерського) рівня освіти.

УДК 004:51-7-047.37(076)

© Хомяк М.Я., 2025

© Волинський національний
університет імені Лесі Українки, 2025

Зміст

Вступ	4
Лабораторна робота №1. Тема: Первинна обробка експериментальних даних психолого-педагогічного дослідження	6
Лабораторна робота №2. Тема: Візуальний аналіз даних психолого-педагогічних досліджень.....	8
Лабораторна робота №3. Тема: Аналіз даних за допомогою зведених таблиць та варіаційної статистики.....	10
Лабораторна робота №4. Тема: Виведення описової статистики у SPSS та MS Excel.....	12
Лабораторна робота №5. Тема: Побудова довірчих інтервалів у статистичному аналізі.....	14
Лабораторна робота №6. Тема: Перевірка нормальності розподілу даних.....	16
Лабораторна робота №7. Тема: Тестування гіпотез про рівність середніх значень	21
Лабораторна робота №8. Тема: Дослідження кореляції.....	24
Лабораторна робота №9. Тема: Побудова моделі лінійної регресії.....	27

ВСТУП

У сучасній освіті зростає необхідність у використанні математичних методів для аналізу даних, особливо у сфері психолого-педагогічних досліджень. Обробка та інтерпретація експериментальних даних є важливими складовими наукової діяльності, що дозволяють отримати обґрунтовані висновки та ухвалювати ефективні рішення в освітньому процесі.

Даний лабораторний практикум спрямований на формування у студентів практичних навичок роботи з основними статистичними методами, які застосовуються в аналізі психолого-педагогічних досліджень. В ході виконання лабораторних робіт здобувачі освіти ознайомляться з первинною обробкою експериментальних даних, методами візуалізації, використанням зведених таблиць, описової статистики, тестуванням статистичних гіпотез та побудовою регресійних моделей.

Особливістю даного практикуму є використання сучасних інструментів обробки даних, таких як SPSS Statistics, MS Excel, Power BI, Python та R, що дає змогу студентам здобути навички роботи з професійним програмним забезпеченням. Завдання сформульовані таким чином, щоб студенти могли не лише отримати теоретичні знання, а й навчитися застосовувати їх на практиці.

Лабораторний практикум буде корисним для здобувачів другого (магістерського) рівня освіти спеціальності **014 Середня освіта (Інформатика)** та сприятиме формуванню компетентностей у сфері обробки та аналізу освітніх даних.

Практикум містить 9 лабораторних робіт, які охоплюють основні методи математичного аналізу даних у психолого-педагогічних дослідженнях:

1. **Первинна обробка експериментальних даних** – закладає основи підготовки даних до аналізу.
2. **Візуальний аналіз даних** – знайомить студентів із графічними методами представлення інформації.

3. **Зведені таблиці та варіаційна статистика** – демонструє підходи до організації та узагальнення даних.
4. **Описова статистика у SPSS та MS Excel** – розглядає сучасні інструменти для статистичного аналізу.
5. **Побудова довірчих інтервалів** – вивчає методи оцінювання параметрів вибірки.
6. **Перевірка нормальності розподілу** – аналізує критерії та методи визначення розподілу даних.
7. **Тестування гіпотез про рівність середніх значень** – знайомить студентів із методами порівняння вибірок.
8. **Дослідження кореляції** – розглядає методи виявлення залежностей між змінними.
9. **Побудова моделі лінійної регресії** – дозволяє студентам зрозуміти основи прогнозування на основі статистичних даних.

Практикум містить детальні інструкції, які сприяють формуванню у студентів практичних навичок роботи зі статистичними методами. Значною перевагою є використання сучасного програмного забезпечення, зокрема SPSS Statistics, MS Excel, Power BI, Python та R, що відповідає актуальним вимогам до підготовки фахівців у галузі освіти та аналізу даних.

Лабораторна робота №1.

Тема: Первинна обробка експериментальних даних психолого-педагогічного дослідження

Мета завдання: Ознайомлення з методами первинної обробки даних, одержаних в експериментальних дослідженнях

Завдання:

1. Імпорт даних

- Завантажте наданий набір даних у MS Excel (або інший інструмент за вашим вибором, наприклад, Python із бібліотекою pandas або R).
- Переконайтеся, що стовпці правильно розпізнані та немає помилок у форматуванні.

2. Перевірка даних

Перевірте набір даних на наявність пропусків (у цьому синтетичному наборі їх немає, але виконайте цю процедуру як тренування для реальних даних).

3. Обчислення основних статистичних характеристик

Для кожного показника (успішність, стрес, залученість, час на навчання, задоволення) обчисліть:

- Середнє арифметичне (Mean)
- Медіану (Median)
- Стандартне відхилення (Standard Deviation)
- Мінімальне та максимальне значення (Min, Max)
- У MS Excel використовуйте функції: AVERAGE, MEDIAN, MODE, STDEV, MIN, MAX. У Python – pandas (describe()), у R – summary() або відповідні функції.

4. Звіт

Складіть короткий звіт із результатами:

- Які значення основних статистичних характеристик ви отримали для кожного показника?
- Які попередні висновки можна зробити про дані (наприклад, чи є значна варіативність у часі на навчання, чи високий середній рівень стресу)?

Інструменти:

- Python (бібліотека pandas для обчислень)
- R (базові функції або пакет dplyr)
- Power BI або Tableau (якщо доступні)
- SPSS Statistics
- Excel

Форма звіту

- Excel-файл (або Google Sheets), що містить вихідні та оброблені дані, результати обчислень, зведені таблиці.
- Текстовий документ із коротким описом методів обробки та висновками.
- Скріншоти із поясненнями.

Лабораторна робота №2.

Тема: Візуальний аналіз даних психолого-педагогічних досліджень

Мета завдання: Ознайомлення із методами візуалізації даних, розробка інтерактивного дашборду для аналізу результатів психолого-педагогічного дослідження

Завдання:

1. Обробка даних

Проведіть попередню обробку даних: перевірте на пропуски, визначте основні статистичні характеристики (середнє, медіану, моду, стандартне відхилення), можна використати результати виконання лабораторної роботи №1.

2. Візуалізація даних

Побудуйте відповідні діаграми гістограми, стовпчикові, кругові чи коробкові діаграми

3. Створення інтерактивного дашборду

Використовуйте інструменти для візуалізації даних, такі як Python (бібліотеки Dash, Plotly, Matplotlib, Seaborn), R (Shiny), Power BI або Tableau, Excel, SPSS. Спробуйте різні інструменти аби порівняти результати. Додайте динамічні графіки та таблиці.

4. Інтерпретація результатів

Проаналізувати отримані візуалізації, зробити висновки щодо досліджуваної проблеми.

Запропонувати можливі педагогічні рекомендації на основі отриманих результатів.

Інструменти:

- Python (Jupyter Notebook, Dash, Plotly)
- R (RStudio, Shiny)

- Power BI або Tableau (якщо доступні)
- SPSS Statistics
- Excel

Форма звіту

- Код або файл дашборду.
- Опис підготовки даних та реалізації візуалізації.
- Висновки та інтерпретація результатів.

Лабораторна робота №3.

Тема: Аналіз даних за допомогою зведених таблиць та варіаційної статистики

Мета: Ознайомлення із методами використання зведених таблиць для аналізу психолого-педагогічних даних

Завдання

1. Завантаження та підготовка даних

- Завантажити наданий файл із даними.
- Визначити основні змінні для аналізу.

2. Створення зведених таблиць у MS Excel (або Google Sheets, SPSS, Python)

- Створити зведену таблицю, яка відображає середнє значення успішності залежно від рівня стресу.
- Додати до таблиці підрахунок кількості студентів у кожній категорії стресу.
- Визначити середній час навчання для студентів із різним рівнем залученості.
- Побудувати частотний розподіл рівня задоволення навчанням.

3. Розрахунок характеристик варіаційної статистики

- Визначити **середнє значення, медіану, моду, мінімальне та максимальне значення** для показників успішності, стресу, залученості, часу на навчання, задоволення.
- Обчислити **дисперсію, стандартне відхилення та коефіцієнт варіації** для цих змінних.
- Проаналізувати розподіл успішності: визначити асиметрію (зсув вправо/вліво) та ексцес (гостроверхий чи плоский розподіл).

4. Інтерпретація результатів

- Визначити, чи є значні відмінності у варіації рівня успішності та задоволеності навчанням серед студентів із різними рівнями стресу.

- Зробити висновки щодо варіативності показників та їх стійкості (які змінні мають найбільше розсіювання, які — найбільш стабільні).

Форма звіту

- Файл Excel (або Google Sheets) із зведеними таблицями та розрахунками.
- Короткий текстовий звіт із висновками (опис отриманих закономірностей).

Лабораторна робота №4.

Тема: Виведення описової статистики у SPSS та MS Excel

Мета завдання: Ознайомитись із методами обчислення описової статистики у SPSS та MS Excel, зрозуміти відмінності між Excel і SPSS у роботі з даними

Завдання:

1. Підготовка даних

1. Завантажити синтетичний набір даних (успішність, стрес, залученість, час на навчання, задоволення навчанням).
2. Ознайомитися зі структурою даних, перевірити коректність введення.

2. Виведення описової статистики у MS Excel

1. Використання пакета аналізу «Описова статистика» (Descriptive Statistics)
 - Увімкнути надбудову **Пакет аналізу (Data Analysis)** в MS Excel.
 - Вибрати «Описова статистика» (**Descriptive Statistics**) і виконати аналіз для всіх змінних.
 - Відзначити наступні характеристики:
 - Середнє значення (Mean)
 - Медіана (Median)
 - Мода (Mode)
 - Стандартне відхилення (Standard deviation)
 - Діапазон (Range: max - min)
 - Асиметрія (Skewness)
 - Екссес (Kurtosis)
 - Аналіз отриманих результатів
 - Визначити змінну з найбільшою варіативністю.
 - Оцінити розподіл даних (нормальний чи асиметричний).

3. Виведення описової статистики у SPSS

1. Імпорт даних у SPSS

- Відкрити SPSS та завантажити дані (можна скопіювати їх з Excel або відкрити **.xlsx** чи **.csv** файл).

2. Обчислення описової статистики

- Використати вкладку **Analyze** → **Descriptive Statistics** → **Descriptives...**
- Вибрати змінні для аналізу та увімкнути параметри:
 - **Mean** (середнє значення)
 - **Std. Deviation** (стандартне відхилення)
 - **Variance** (дисперсія)
 - **Range** (розмах)
 - **Minimum, Maximum**

3. Розширений аналіз через Frequencies

- Використати **Analyze** → **Descriptive Statistics** → **Frequencies...**
- Увімкнути опції **Skewness** (асиметрія) та **Kurtosis** (ексцес) для оцінки розподілу змінних.

4. Інтерпретація отриманих результатів

1. Порівняти результати аналізу в SPSS та Excel.
2. Виявити закономірності у даних (наприклад, чи є змінні з великою варіативністю, чи є асиметрія в розподілі).
3. Зробити висновки про особливості успішності, стресу, залученості студентів.

Форма звіту

- Файл Excel із розрахунками описової статистики.
- Звіт SPSS у форматі **.sav** або **.spv**.
- Короткий текстовий звіт із висновками (опис результатів аналізу та порівняння методів Excel і SPSS).

Лабораторна робота №5.

Тема: Побудова довірчих інтервалів у статистичному аналізі

Мета

- Навчитися будувати довірчі інтервали для середнього значення на основі реального набору даних у SPSS та MS Excel.
- Інтерпретувати отримані результати.

Завдання

1. Підготовка даних

1. Відкрити **SPSS** та завантажити файл **опитування_CO_Інф_внутрішнє.sav**.
2. Ознайомитися з доступними змінними у файлі.
3. Вибрати 2-3 ключові змінні для аналізу (наприклад, **успішність**, **рівень стресу**, **залученість у навчання**).

2. Побудова довірчих інтервалів у SPSS

1. Використання аналізу середнього значення

- Відкрити вкладку **Analyze** → **Compare Means** → **One-Sample T Test**.
- Обрати змінну для аналізу (наприклад, **успішність**).
- В полі **Test Value** залишити **0**, оскільки нам потрібен лише довірчий інтервал.
- Встановити рівень довіри **95%** (можна також спробувати **99%**).
- Натиснути **OK** та переглянути отримані довірчі інтервали у вихідних даних (**Confidence Interval for the Mean**).

2. Додатковий аналіз через Descriptives

- **Analyze** → **Descriptive Statistics** → **Explore**.
- Обрати змінну (наприклад, **залученість у навчання**).
- У вкладці **Statistics** увімкнути **Confidence Interval for Mean** та обрати **95%**.
- Натиснути **OK** та переглянути результати у вихідному вікні.

3. Побудова довірчих інтервалів у MS Excel

1. Експорт даних з SPSS у Excel

- **File** → **Export** → **Excel** (зберегти дані у форматі **.xlsx**).

2. Розрахунок довірчого інтервалу вручну у Excel

- Відкрити збережений **.xlsx** файл.
- Використати формулу для довірчого інтервалу: $CI = \bar{X} \pm t_{\alpha/2} \cdot s_n$
- Використати наступні формули у Excel:
 - **=AVERAGE(range)** – середнє значення
 - **=STDEV.S(range)** – стандартне відхилення
 - **=COUNT(range)** – кількість значень
 - **=T.INV.2T(0.05, n-1)** – критичне значення t-розподілу

3. Автоматичне обчислення через «Аналіз даних»

- Включити **Пакет аналізу (Data Analysis)**.
- Використати **Descriptive Statistics (Описова статистика)** та увімкнути **Confidence Level for Mean (Рівень довіри для середнього)**.

4. Аналіз отриманих результатів

1. Порівняти довірчі інтервали для різних змінних у SPSS та Excel.
2. Проаналізувати, чи потрапляє певне теоретичне очікуване значення в отриманий довірчий інтервал.
3. Оцінити, для яких змінних довірчий інтервал ширший або вужчий (від чого це залежить?).

Форма звіту

- **Файл SPSS (.sav)** із виконаними аналізами.
- **Файл Excel (.xlsx)** із розрахунками довірчих інтервалів.
- **Короткий текстовий звіт (Word/PDF)** із висновками щодо отриманих довірчих інтервалів.

Лабораторна робота №6.

Тема: Перевірка нормальності розподілу даних

Мета:

- Перевірити гіпотезу щодо нормального закону розподілу досліджуваних ознак.
- Використати SPSS, MS Excel, Python або R для аналізу.
- Інтерпретувати отримані результати та зробити висновки.

Вихідні дані: Файл "опитування_кібер.sav", що містить дані.

Завдання:

1. Формулювання гіпотез:

- **H_0 (нульова гіпотеза):** Дані для кожного показника мають нормальний розподіл.
- **H_1 (альтернативна гіпотеза):** Дані не відповідають нормальному розподілу.

2. Імпорт даних:

- **SPSS:** File → Open → Data, вибрати файл "опитування_кібер.sav".
- **Python:**
 - `import pandas as pd`
 - `import pyreadstat`
 - `df, meta = pyreadstat.read_sav("опитування_кібер.sav")`
 - `print(df.head())`
- **R:**
 - `library(haven)`
 - `df <- read_sav("опитування_кібер.sav")`
 - `head(df)`
- **MS Excel:** Додаткове збереження у CSV з SPSS або використання інших методів конвертації.

3. Перевірка нормальності розподілу:

3.1. Візуальна оцінка нормальності

Використовуємо графічні методи для оцінки нормальності розподілу.

1) Гістограма (Histogram)

Дозволяє оцінити форму розподілу та порівняти її з нормальним розподілом.

- **SPSS:** Graphs → Legacy Dialogs → Histogram, вибрати змінну та позначити "Display normal curve".
- **MS Excel:** Використати Гістограму (Вставка → Гістограма).
- **Python:**
 - `import matplotlib.pyplot as plt`
 - `import seaborn as sns`
 - `sns.histplot(df["Змінна"], kde=True, bins=10)` # `kde=True` додає криву щільності
 - `plt.show()`
- **R:**
 - `hist(df$Змінна, probability=TRUE, col="lightblue", main="Гістограма з кривою нормального розподілу")`
 - `lines(density(df$Змінна), col="red", lwd=2)`

☐ Інтерпретація:

- Якщо гістограма має симетричну дзвоноподібну форму, розподіл близький до нормального.
- Якщо є зміщення (скошеність) або ексцес, розподіл може бути не нормальним.
-

2) Q-Q (Квантиль-квантиль) графік

Порівнює квантилі досліджуваних даних із квантилями нормального розподілу.

- **SPSS:** Analyze → Descriptive Statistics → Explore → Plots → Normal Q-Q plot.
- **MS Excel:** Потрібен ручний розрахунок квантилів.

- Python:
- `import scipy.stats as stats`
- `import matplotlib.pyplot as plt`
- `stats.probplot(df["Змінна"], dist="norm", plot=plt)`
`plt.show()`
- R:
- `qqnorm(df$Змінна)`
`qqline(df$Змінна, col="red")`

☛☐ Інтерпретація:

- Якщо точки лежать на прямій лінії, розподіл нормальний.
- Якщо є вигини або відхилення, розподіл може бути не нормальним.

3) Boxplot (Ящик з вусами)

Оцінює розмах даних, викиди та симетрію розподілу.

- SPSS: Graphs → Boxplot → Simple.
- MS Excel: Вставка → Діаграма → Коробковий графік (Box and Whisker).
- Python:
- `sns.boxplot(y=df["Змінна"])`
`plt.show()`
- R:
- `boxplot(df$Змінна, main="Boxplot розподілу", col="lightblue")`

☛☐ Інтерпретація:

- Якщо медіана по центру, а "вуса" рівномірні → нормальний розподіл.
- Якщо "вуса" асиметричні або є багато викидів → можливе відхилення від нормальності.

3.2. Тести на нормальність:

- Тест Колмогорова-Смирнова
- Тест Шапіро-Уїлка

Методи розрахунку:

- SPSS: Analyze → Descriptive Statistics → Explore → Plots → Normality tests.
- Python:
 - from scipy.stats import shapiro, kstest
 - for col in df.columns:
 - stat, p = shapiro(df[col])
 - print(f"{col}: Shapiro-Wilk test p-value = {p:.4f}")
- R:
 - for (col in colnames(df)) {
 - print(shapiro.test(df[[col]]))
 - }
- MS Excel: Обмежені можливості, потрібно розраховувати вручну або використовувати надбудови.

4. Інтерпретація результатів:

- $p\text{-value} > 0.05 \rightarrow$ не відхиляємо H_0 (розподіл нормальний).
- $p\text{-value} \leq 0.05 \rightarrow$ відхиляємо H_0 (розподіл не нормальний).
- Аналізуємо асиметрію та ексцес, щоб підтвердити або спростувати результати тестів.

Для коректної оцінки нормальності поєднуємо статистичні тести та графічні методи.

Якщо всі методи підтверджують нормальність – можна застосовувати параметричні методи (наприклад, t-тест).

Якщо є відхилення – краще використовувати непараметричні методи аналізу (наприклад, тест Манна-Уїтні).

5. Звіт

- Таблиця описової статистики.
- Результати тестів на нормальність.
- Висновки щодо нормального розподілу.

Форма подання:

- Файл SPSS (.sav) або .spv з результатами.
- Файл Excel/Python/R із розрахунками.
- Текстовий документ із висновками.

Лабораторна робота №7.

Тема: Тестування гіпотез про рівність середніх значень

Мета роботи: Перевірити гіпотези про рівність середніх значень для досліджуваних змінних у файлі **опитування_кібер_2.sav** за допомогою статистичних методів у **SPSS, MS Excel, Python або R**.

Завдання

1. Формулювання гіпотез

- **Нульова гіпотеза (H_0):** Середні значення досліджуваних змінних для двох груп (наприклад, за статтю, рівнем стресу тощо) не відрізняються.
- **Альтернативна гіпотеза (H_1):** Існує статистично значуща різниця між середніми значеннями у вибраних групах.

2. Імпорт та ознайомлення з даними

- Завантажте файл **опитування_кібер_2.sav** у **SPSS** (**File** → **Open** → **Data**).
- Використайте **Variable View**, щоб переглянути змінні у файлі.
- Визначте дві групи, між якими порівнюватимуться середні значення (наприклад, чоловіки та жінки, рівень стресу: високий / низький).

3. Обчислення описової статистики

- Обчисліть **середнє значення, стандартне відхилення** для кожної групи.
- У **SPSS**: **Analyze** → **Compare Means** → **Means**.
- У **Excel**: **AVERAGE()** для кожної групи.
- У **Python**:
 - `import pandas as pd`
 - `df = pd.read_spss("опитування_кібер_2.sav")`
 - `df.groupby("Групова_змінна")["Змінна"].describe()`
- У **R**:

- `df <- haven::read_sav("опитування_кібер_2.sav")`
- `aggregate(Змінна ~ Група, data=df, summary)`

4. Перевірка передумов для t-тесту

- **Перевірка нормальності розподілу**
 - Використовуйте тест Шапіро-Уїлка (**Analyze → Descriptive Statistics → Explore → Normality**).
- **Перевірка рівності дисперсій (тест Левена)**
 - У SPSS: **Analyze → Compare Means → One-Way ANOVA → Homogeneity of Variances**.
 - У Python:
 - `from scipy.stats import levene`
 - `stat, p = levene(df[df["Група"]=="A"]["Змінна"], df[df["Група"]=="B"]["Змінна"])`
 - `print(f"Levene Test: p-value = {p}")`

5. Тестування гіпотез про рівність середніх значень

- Якщо дані **нормально розподілені та дисперсії рівні** → Використовуємо **t-тест**.
- Якщо **дисперсії нерівні** → t-тест з урахуванням нерівності дисперсій (**Welch's t-test**).
- Якщо дані **не нормальні** → Використовуємо **непараметричний тест Манна-Уїтні**.

У SPSS:

- **Analyze → Compare Means → Independent-Samples T Test**
- Виберіть **залежну змінну** та **групову змінну** (наприклад, стать).
- Натисніть "Define groups" та вкажіть значення (наприклад, 1 = чоловіки, 2 = жінки).
- Перевірте p-value: якщо **$p \leq 0.05$** , різниця значуща.

У Python (t-тест):

`from scipy.stats import ttest_ind`

```
stat, p = ttest_ind(df[df["Група"]=="A"]["Змінна"],
df[df["Група"]=="B"]["Змінна"], equal_var=True)
print(f"T-test: p-value = {p}")
```

У Python (Манна-Уїтні, якщо розподіл не нормальний):

```
from scipy.stats import mannwhitneyu
stat, p = mannwhitneyu(df[df["Група"]=="A"]["Змінна"],
df[df["Група"]=="B"]["Змінна"])
print(f"Mann-Whitney U Test: p-value = {p}")
```

6. Інтерпретація результатів

- Якщо **p-value** ≤ 0.05 → Відхиляємо H_0 (середні значення відрізняються).
- Якщо **p-value** > 0.05 → Не відхиляємо H_0 (немає доказів значущої різниці).
- Додайте **візуалізацію (гістограми, boxplot)** для порівняння груп.

7. Звіт

- Коротко опишіть досліджувані змінні.
- Наведіть середні значення та стандартні відхилення.
- Опишіть результати перевірки нормальності.
- Вкажіть результати t-тесту або Манна-Уїтні.
- Напишіть висновок: чи є значуща різниця між групами?

Лабораторна робота №8.

Тема: Дослідження кореляції

Мета роботи: Визначити ступінь зв'язку між досліджуваними змінними за допомогою коефіцієнта кореляції **Пірсона** та пояснити тісноту зв'язку.

Вихідні дані: Файл опитування_CO_мат.sav, що містить дані дослідження.

Завдання

1. Формулювання гіпотез:

- **Нульова гіпотеза (H_0):** Кореляція між змінними відсутня (коефіцієнт кореляції Пірсона ≈ 0).
- **Альтернативна гіпотеза (H_1):** Існує статистично значущий кореляційний зв'язок між змінними (коефіцієнт кореляції Пірсона $\neq 0$).

2. Імпорт та ознайомлення з даними:

- Завантажте файл **опитування_кібер_2.sav** у **SPSS** (**File** → **Open** → **Data**).
- Використайте **Variable View**, щоб переглянути доступні змінні.
- Оберіть дві кількісні змінні, між якими перевірятиметься кореляція (наприклад, "час онлайн" і "рівень стресу").

3. Перевірка нормальності розподілу змінних:

- Використовуйте **тест Шапіро-Уїлка** для перевірки нормальності (**Analyze** → **Descriptive Statistics** → **Explore** → **Normality**).
- Якщо обидві змінні мають нормальний розподіл → застосовуємо **кореляцію Пірсона**.
- Якщо хоча б одна змінна **не** має нормального розподілу → застосовуємо **кореляцію Спірмена**.

У SPSS:

- **Analyze → Correlate → Bivariate.**
- Виберіть дві змінні для аналізу.
- Увімкніть **Pearson** та опцію "Two-tailed significance".
- Натисніть ОК – отримаєте таблицю з кореляціями та значенням p-value.

У Python:

```
import pandas as pd
```

```
from scipy.stats import pearsonr
```

```
# Завантаження даних
```

```
df = pd.read_spss("опитування_CO_мат.sav")
```

```
# Розрахунок коефіцієнта кореляції Пірсона
```

```
corr, p_value = pearsonr(df["Час_онлайн"], df["Рівень_стресу"])
```

```
print(f'Кореляція Пірсона: {corr:.3f}, p-value: {p_value:.3f}')
```

У R:

```
df <- haven::read_sav("опитування_CO_мат.sav")
```

```
cor.test(df$Час_онлайн, df$Рівень_стресу, method="pearson")
```

5. Інтерпретація результатів:

- Якщо **p-value** ≤ **0.05**, відхиляємо H_0 – зв'язок між змінними є статистично значущим.
- Якщо **p-value** > **0.05**, H_0 не відхиляється – кореляція статистично незначуща.
- **Оцінка тісноти зв'язку:**
 - $|r| < 0.3$ → слабкий зв'язок
 - $0.3 \leq |r| < 0.5$ → середній зв'язок
 - $|r| \geq 0.5$ → сильний зв'язок

6. Візуалізація кореляції (за бажанням)

- Побудуйте **розсіяння (scatter plot)** для візуальної оцінки.

У Python:

```
import matplotlib.pyplot as plt
```

```
import seaborn as sns
```

```
sns.scatterplot(x=df["Час_онлайн"], y=df["Рівень_стресу"])
```

```
plt.xlabel("Час онлайн (години)")
```

```
plt.ylabel("Рівень стресу")
```

```
plt.title("Залежність рівня стресу від часу онлайн")
```

```
plt.show()
```

У SPSS:

- **Graphs → Chart Builder → Scatter/Dot.**
- Виберіть "Simple Scatter" і перетягніть змінні.

7. Звіт

- Коротко опишіть змінні.
- Наведіть коефіцієнт кореляції та p-value.
- Поясніть, чи є статистично значущий зв'язок між змінними.
- Висновок: наскільки сильний зв'язок та його практичне значення.

Лабораторна робота №9.

Тема: Побудова моделі лінійної регресії

Мета роботи: Побудувати модель лінійної регресії для виявлення залежності між двома змінними проаналізувати значущість моделі та зробити висновки

Завдання

1. Формулювання гіпотез:

- **Нульова гіпотеза (H_0):** Незалежна змінна не впливає на залежну (коефіцієнт регресії = 0).
- **Альтернативна гіпотеза (H_1):** Існує статистично значущий лінійний зв'язок між змінними (коефіцієнт регресії $\neq 0$).

2. Імпорт та ознайомлення з даними:

- Завантажте файл **опитування_кібер_2.sav** у **SPSS** (**File** → **Open** → **Data**).
- Перегляньте змінні у **Variable View**.
- Оберіть одну змінну як **залежну (Y)** (наприклад, рівень стресу) та одну як **незалежну (X)** (наприклад, час онлайн).

3. Перевірка кореляції між змінними:

- Використайте коефіцієнт **Пірсона** (**Analyze** → **Correlate** → **Bivariate**).
- Якщо кореляція слабка ($|r| < 0.3$), лінійна регресія може бути неефективною.

4. Побудова моделі лінійної регресії у SPSS:

- **Analyze** → **Regression** → **Linear**.
- Виберіть залежну змінну (наприклад, рівень стресу).
- Виберіть незалежну змінну (наприклад, час онлайн).
- Натисніть **ОК** → отримаєте таблиці з параметрами моделі.

У Python:

```
import pandas as pd
```

```
import statsmodels.api as sm
```

```
# Завантаження даних
```

```
df = pd.read_spss("опитування_кібер_2.sav")
```

```
# Визначення змінних
```

```
X = df["Час_онлайн"]
```

```
Y = df["Рівень_стресу"]
```

```
# Додавання константи
```

```
X = sm.add_constant(X)
```

```
# Побудова моделі
```

```
model = sm.OLS(Y, X).fit()
```

```
print(model.summary())
```

У R:

```
df <- haven::read_sav("опитування_кібер_2.sav")
```

```
model <- lm(Рівень_стресу ~ Час_онлайн, data=df)
```

```
summary(model)
```

5. Аналіз параметрів моделі:

- **Коефіцієнт детермінації (R^2):** показує, яка частка варіації Y пояснюється X .
- **Коефіцієнт регресії (β):** показує, наскільки змінюється Y при зміні X на 1 одиницю.
- **p-value:** якщо ≤ 0.05 , відхиляємо H_0 , тобто X значущо впливає на Y .

6. Візуалізація регресійної моделі

У Python:

```
import matplotlib.pyplot as plt
```

```
import seaborn as sns
```

```
sns.regplot(x=df["Час_онлайн"], y=df["Рівень_стресу"])  
plt.xlabel("Час онлайн (год)")  
plt.ylabel("Рівень стресу")  
plt.title("Лінійна регресія: Час онлайн → Рівень стресу")  
plt.show()
```

У SPSS:

- **Graphs → Chart Builder → Scatter/Dot.**
- Додайте трендову лінію регресії.

7. Звіт

- Опишіть змінні.
- Наведіть рівняння регресії ($Y = \beta_0 + \beta_1 X$).
- Проаналізуйте значущість моделі (p-value, R^2).
- Візуалізуйте дані та зробіть висновки.

РЕКОМЕНДОВАНА ЛІТЕРАТУРА

Методичне забезпечення

1. Maria Khomyak & Svitlana Yatsyuk. SOCIOMETRIC TEST AS A METHOD OF STUDYING FOR THE INDIVIDUAL AND THE TEAM. PROCESSING OF RESEARCH RESULTS AND USING THEM IN THE EDUCATIONAL PROCESS. *Proceeding Book of III International Congress of modern studies in social sciences and humanities*, June 7-10, 2022, Karabagh, Azerbaijan. Vol.3, P. 58-62.

2. Khomyak M. Analysis of data on the organization of distance learning. *Middle East International conference on contemporary scientific studies-V*, March 27-28, 2021, Ankara, Turkey. Vol.II. P. 384-386.

3. Хомяк М.Я. Дистанційний курс Moodle: Математичні методи обробки даних психолого-педагогічних досліджень (в тестовому режимі). Режим доступу: <https://moodle.vnu.edu.ua/course/view.php?id=>

4. Хомяк М.Я. Мова програмування R як засіб навчання математичної статистики майбутніх ІТ-фахівців та вчителів інформатики. *Математика. Інформаційні технології. Освіта*: тези доп. XI міжнар. наук.-практ. конф. (м. Луцьк, 3-5 червн. 2022 р.). Луцьк, 2022.. С. 171-173.

5. Хомяк М. Я. Математичні методи аналізу емпіричних даних: методичні вказівки. Луцьк: ВНУ ім. Лесі Українки, 2021. 16 с.

6. Хомяк М. Я. Математичні методи обробки даних психолого-педагогічних досліджень: методичні вказівки. Луцьк : ВНУ ім. Лесі Українки, 2022. 16 с.

7. Яцюк С.М., Хомяк М.Я., Юнчик В.Л., Чепрасова Т.І. Методика використання цифрових освітніх ресурсів у процесі підготовки майбутніх учителів інформатики. *Професіоналізм педагога: теоретичні й методичні аспекти*. 2021. №16. С.15-25.

Рекомендовані джерела

1. Breheny M., Horrell, B. & Stephens, C. (2020). A participatory

journal/dialogue approach to narrative research illustrated using a study of informal caregiving. Qualitative Research in Psychology, DOI: 10.1080/14780887.2020.1716423

2. Маханець Л.Л., Вінничук О.Ю., Григорків М.В. Статистика: лабораторний практикум у STATISTICA 12: навч. посіб. Чернівці : Чернівець. нац. ун-т ім. Ю. Федьковича, 2023. 161 с.

3. Петровська І.Р., Островська К.О. Математично-статистичні методи обробки емпіричних даних психолого-педагогічних досліджень: навчальний посібник. Львів: Друкарня «Справи Кольпінга в Україні», 2021. 140 с

4. Технології оброблення великих даних: конспект лекцій з дисципліни «Технології оброблення великих даних» [Електронний ресурс] : навч. посіб. для студ. спеціальності 121 «Інженерія програмного забезпечення» (освітня програма «Інженерія програмного забезпечення мультимедійних та інформаційно-пошукових систем»)/ Л.М. Олещенко; КПІ ім. Ігоря Сікорського. Електронні текстові дані (1 файл: 5,55 Мбайт). Київ: КПІ ім. Ігоря Сікорського, 2021. 227 с.

Електронне мережне методичне видання

Марія Хомяк

МАТЕМАТИЧНІ МЕТОДИ ОБРОБКИ ДАНИХ
ПСИХОЛОГО-ПЕДАГОГІЧНИХ ДОСЛІДЖЕНЬ

Лабораторний практикум для здобувачів освіти спеціальності

014 Середня освіта (Інформатика)

другого (магістерського) рівня освіти

Друкується в авторській редакції