

ОПРАЦЮВАННЯ СТАТИСТИЧНИХ ДАНИХ ПЕДАГОГІЧНОГО ДОСЛІДЖЕННЯ ЗАСОБАМИ МОВИ R

Джунь Й. В.

доктор фізико-математичних наук, професор,
кафедра математичного моделювання
Приватного вищого навчального закладу «Міжнародний економіко-гуманітарний університет
імені академіка Степана Дем'янчука»
ORCID ID: 0000-0002-5706-3454

Ясінський А. М.

кандидат педагогічних наук, доцент,
кафедра математичного моделювання
Приватного вищого навчального закладу «Міжнародний економіко-гуманітарний університет
імені академіка Степана Дем'янчука»
ORCID ID: 0000-0002-1894-1314

Соловей Л. Я.

старший викладач,
кафедра інформаційних систем та обчислювальних методів
Приватного вищого навчального закладу «Міжнародний економіко-гуманітарний університет
імені академіка Степана Дем'янчука»
ORCID ID: 0009-0001-2832-1741

У педагогічних дослідженнях особливу роль відіграє статистичний аналіз, оскільки саме він дає можливість оцінити ефективність методичних підходів, виявити кореляційні зв'язки між освітніми факторами та підтвердити або спростувати гіпотези. Застосування сучасних інструментів аналізу даних сприяє підвищенню точності досліджень, забезпечує автоматизацію обчислювальних процесів і покращує візуалізацію результатів. Одним із потужних засобів для обробки статистичних даних у наукових дослідженнях є мова програмування R, яка пропонує широкий спектр функцій для математичного моделювання, обробки даних, тестування гіпотез і побудови графіків.

У статті розглянуто особливості використання мови R для опрацювання статистичних даних у педагогічних дослідженнях. Проаналізовано основні пакети та функції R, що забезпечують обробку, візуалізацію та інтерпретацію результатів. Описано методи описової статистики, кореляційного та регресійного аналізу, а також їхню роль у підтвердженні наукових гіпотез. Наведено приклади застосування R у педагогічній сфері, що демонструють ефективність і доступність цього інструменту для дослідників. У дослідженні розроблено методіку проведення аналізу результатів педагогічного дослідження в середовищі RStudio, а також продемонстровано основні методи обробки даних, візуалізації та інтерпретації отриманих результатів.

У роботі детально проаналізовано методи попередньої обробки даних, нормалізації, описової статистики, кореляційного та регресійного аналізу, а також тестування статистичних гіпотез, зокрема t-тест, χ^2 -тест та дисперсійний аналіз (ANOVA).

Запропонований підхід можуть використовувати дослідники в галузі педагогіки, які працюють з емпіричними даними та прагнуть використовувати сучасні інструменти аналізу. Використання R забезпечує високу гнучкість, автоматизацію обчислювальних процесів і відтворюваність результатів, що є актуальним для наукових досліджень.

Ключові слова. статистичний аналіз, педагогічне дослідження, мова R, опрацювання даних, візуалізація, кореляція, регресія.

Dzhun J. V., Yasinsky A. M., Solovei L. Ya. Processing statistical data of pedagogical research using the R language

In pedagogical research, statistical analysis plays a special role, since it allows you to assess the effectiveness of methodological approaches, identify correlations between educational factors, and confirm or refute hypotheses. The use of modern data analysis tools helps to increase the accuracy of research, automates computational processes, and improves the visualization of results. One of the powerful tools for processing

statistical data in scientific research is the R programming language, which offers a wide range of functions for mathematical modeling, data processing, hypothesis testing, and graphing.

The article examines the features of using the R language to process statistical data in pedagogical research. The main R packages and functions that provide processing, visualization, and interpretation of results are analyzed. Methods of descriptive statistics, correlation, and regression analysis are described, as well as their role in confirming scientific hypotheses. Examples of the use of R in the pedagogical field are given, demonstrating the effectiveness and accessibility of this tool for researchers. The study developed a methodology for analyzing the results of pedagogical research in the RStudio environment, and demonstrated the main methods of data processing, visualization and interpretation of the obtained results.

The work analyzed in detail the methods of data preprocessing, normalization, descriptive statistics, correlation and regression analysis, as well as statistical hypothesis testing, in particular the t-test, χ^2 -test and analysis of variance (ANOVA).

The proposed approach can be used by researchers in the field of pedagogy who work with empirical data and seek to use modern analysis tools. The use of R provides high flexibility, automation of computational processes and reproducibility of results, which is relevant for scientific research.

Key words: *statistical analysis, pedagogical research, R language, data processing, visualization, correlation, regression.*

Вступ. У сучасних педагогічних дослідженнях важливу роль відіграє аналіз статистичних даних, оскільки він дає можливість виявляти закономірності, оцінювати ефективність освітніх методик і робити обґрунтовані висновки. Використання сучасних програмних засобів для обробки даних підвищує точність і надійність результатів. Одним із потужних інструментів для статистичного аналізу є мова програмування R та інтегроване середовище RStudio, які забезпечують широкий набір методів для обробки, візуалізації та інтерпретації даних.

Актуальність дослідження зумовлена необхідністю розробки доступної та ефективної методики аналізу педагогічних досліджень у середовищі RStudio. Хоча мова R широко використовується в економіці, біології та соціології, її потенціал у сфері педагогічних досліджень залишається недостатньо розкритим.

Метою цього дослідження є розробка методики проведення аналізу результатів педагогічного дослідження в середовищі RStudio.

«Статистичною гіпотезою називається будь-яке припущення щодо виду або параметрів невідомого закону розподілу. У конкретній ситуації статистичну гіпотезу формують як припущення на певному рівні статистичної значущості про властивості генеральної сукупності за оцінками вибірки» [5, с. 195].

Науково-педагогічне дослідження має спиратися на перевірені факти, що підлягають емпіричній верифікації. «Якими б різноманітними не були цілі та методи педагогічних досліджень, отримані в їхньому підсумку

дані завжди є результатами вимірювань різних педагогічних явищ (процесів, станів, властивостей). Під вимірюваннями зазвичай розуміють процедуру приписування чисел об'єктам вивчення відповідно до певних правил» [2, с. 202]. Висновки, до яких приходять науковець під час аналізу даних, спираючись на початкове уявлення про зв'язки між досліджуваними явищами, не вважаються достовірними, якщо не підтверджені методами математичної статистики.

Проблему формування навичок опрацювання статистичних даних педагогічних досліджень детально досліджено в роботах С. Архипової, П. Воловика, С. Гончаренка, В. Жигаря, В. Малихіної, Д. Новикова, М. Салюк, В. Руденко, О. Рудницької та ін., де розкрито сутність різноманітних статистичних методів з аналізу експериментальних даних.

У роботі С. П. Архипової [1, с. 9–14] розглядається застосування статистичних методів для аналізу даних, отриманих у соціально-педагогічних дослідженнях. Авторка акцентує увагу на важливості використання математичної статистики для обґрунтування та підтвердження результатів експериментів у педагогічній сфері.

За оприлюдненими матеріалами педагогічних науці відслідковується тенденція до зростання рівня математизації, у зв'язку із чим експериментальні дослідження супроводжуються різноманітними статистичними показниками. Практично всі обчислювальні роботи виконуються із використанням різноманіт-

них комп'ютерних інструментів. Поширеною практикою є обчислення критеріїв для статистичної оцінки вибірок і перевірки гіпотез. Обчислення критеріїв Пірсона, Фішера, Стюдента з подальшою інтерпретацією отриманих результатів – встановлена практика педагогічного дослідження. Для опрацювання числових даних використовують методики статистичного аналізу, як-от: обчислення частот, середніх значень, групування даних, шкалювання, графічна та діаграмна інтерпретація результатів. Менш поширеними серед науковців є використання складних математично-статистичних методів, а саме регресійного, дисперсійного, дискримінантного, кореляційного аналізу.

У статті [6, с. 453–457] розглядається використання комп'ютерних програм для аналізу даних педагогічних експериментів. Автори аналізують різні статистичні пакети, зокрема Statgraphics, Statistica, S-plus, SPSS та Systat, порівнюючи їх функціональні можливості, зручність використання й ефективність у контексті педагогічних досліджень.

Автори [4] досліджують можливості розповсюджених сучасних статистичних пакетів обробки даних та їх функціональні можливості для використання в науково-дослідницькій діяльності.

З метою обробки статистичних даних може бути розроблена спеціальна комп'ютерна програма призначена для розв'язування задач конкретного дослідження.

У навчально-методичному посібнику [3] розглядається застосування мови R для розв'язання складних завдань математичної статистики та економетрики. Автори надають практичні поради щодо використання R у наукових дослідженнях, що може бути корисним для педагогів і дослідників.

Дослідження [7, с. 121–150] містить основи статистичних методів, як-от вимірювання центральної тенденції, варіативності, кореляції та регресії, з використанням R, що допомагає дослідникам в галузі освіти аналізувати дані та проводити статистичні тести.

Методи та методики дослідження. Орієнтуючись на опрацювання даних у середовищі RStudio для зберігання даних, ми рекомендуємо ознайомитись із структурою CSV файлів і використати їх можливості для зберігання результатів педагогічних вимірювань.

Узагальнену методичну траєкторію процесу опрацювання результатів педагогічного експерименту можна описати такими етапами:

На першому етапі проводяться збір даних і реєстрація результатів експерименту. Це можуть бути оцінки, тести, анкети, опитування тощо. Важливо, щоб ці дані були достовірними та репрезентативними.

Первинна обробка даних зводиться до перевірки правильності введення даних та кодування і класифікації результатів. На цьому етапі також проводять підготовку даних до подальшої статистичної обробки (наприклад, перетворення результатів у числа чи відсотки).

Наступний етап, який зручно виконувати в Rstudio, зводиться до підрахунку основних статистичних показників, як-от середнє значення, медіана, мода, дисперсія, стандартне відхилення, а також визначення розподілу результатів у вигляді гістограми та таблиці частот.

Якщо експеримент передбачає порівняння двох чи більше груп (наприклад, контрольна й експериментальна), то використовуються методи порівняння середніх (наприклад, t-тест для незалежних вибірок). У випадку проведення аналізу більше ніж двох груп застосовуються методи аналізу варіацій (ANOVA).

На етапі кореляційного аналізу, який проводиться в Rstudio, вивчають залежності між різними змінними, використовуючи, наприклад, коефіцієнт кореляції Пірсона. Регресійний аналіз проводять у випадку, якщо потрібно прогнозувати результат на основі інших змінних (лінійна регресія, логістична регресія тощо).

Для підтвердження результатів експерименту використовують критерії значущості, а саме t-критерій, χ -квадрат або інші методи тестування гіпотез.

Результати статистичних обчислень можуть бути представлені у вигляді графіків, діаграм, таблиць, що полегшує сприйняття даних та їх аналіз.

Кінцевий етап опрацювання даних – це інтерпретація результатів. Після проведення статистичних тестів потрібно зробити висновки щодо гіпотез, що перевірялися, та пояснити значення отриманих результатів у контексті педагогічної практики.

Результати досліджень. Важливо врахувати, що статистична обробка повинна бути проведена з урахуванням специфіки педагогіч-

ного експерименту, що може передбачати використання різних методів залежно від розміру вибірки, типу даних і характеру гіпотез. Методологічна траєкторія опрацювання результатів педагогічного експерименту залежить від структури даних, мети експерименту, параметрів вибірки та результатів проміжних обчислень.

Розглянемо поетапно кожен із компонентів процесу обробки даних у контексті педагогічного експерименту в середовищі RStudio.

На етапі збору й підготовки даних потрібно зафіксувати дані, які містять результати експерименту, наприклад оцінки учасників, час виконання завдань, відповіді на анкети тощо. Дані можуть бути зібрані в різних форматах: CSV, Excel, текстові файли. Для роботи з ними в RStudio часто використовують пакети `readr`, `readxl`, або `openxlsx`.

```
library(readr)
data <- read_csv("data.csv")
```

Перед початком аналізу важливо очистити дані від пропущених значень, помилок або аномалій. Для цього використовуються функції з пакетів `dplyr`, `tidyr`, наприклад:

```
library(dplyr)
clean_data <- data %>%
  filter(!is.na(оцінка)) %>%
  mutate(оцінка = as.numeric(оцінка))
t_test_result <- t.test(оцінка ~
  група, data = clean_data)
```

Описова статистика є першим кроком для аналізу даних і допомагає отримати загальну картину досліджуваних змінних. У RStudio є багато вбудованих функцій для обчислення основних статистичних показників, а також для візуалізації даних.

Середнє значення (`mean`) обчислюється за допомогою функції `mean()`. Наприклад, якщо у вас є вектор даних `x`:

```
x <- c(1, 2, 3, 4, 5, 6, 7, 8, 9, 10) #
Приклад даних
mean_x <- mean(x) # Обчислення
середнього значення
mean_x
```

Медіана – це середнє значення в упорядкованому наборі даних. Для обчислення медіани використовується функція `median()`:

```
median_x <- median(x) # Обчислення
медіани
median_x
```

В R немає стандартної функції для обчислення моди, але її можна знайти за допомогою

кастомної функції. Мода – це значення, яке з'являється найчастіше.

```
# Кастомна функція для обчислення моди
get_mode <- function(v)
{uniqv <- unique(v)uniqv[which.max
(tabulate(match(v,uniqv)))]}
mode_x <- get_mode(x) # Обчислення моди
mode_x
```

Для обчислення дисперсії та стандартного відхилення використовуються функції `var()` і `sd()` відповідно.

```
variance_x <- var(x)
variance_x # Дисперсія
sd_x <- sd(x) # Стандартне відхилення
sd_x
```

Таблиця частот показує, скільки разів кожне значення з'являється в наборі даних. В R для цього використовується функція `table()`:

```
# Таблиця частот
frequency_table <- table(x)
frequency_table
```

Функція `summary()` дає загальний огляд даних, зокрема середнє значення, мінімум, максимум, кватилі та медіану.

```
summary(x) # Описова статистика
```

Коробковий графік допомагає візуалізувати розподіл, виявляти аномальні значення та порівнювати кілька груп даних.

```
# Коробковий графік
boxplot(x, main="Коробковий графік",
ylab="Значення", col="lightyellow")
```

Ці методи дають змогу ефективно аналізувати та візуалізувати результати педагогічного експерименту або будь-якого іншого дослідження в RStudio. У разі коли експеримент передбачає порівняння результатів двох або більше груп, статистичні методи дають можливість перевірити, чи є суттєві відмінності між групами. Найпоширенішими методами є t-тест для незалежних вибірок для порівняння двох груп і ANOVA (аналіз варіацій) для порівняння більше ніж двох груп.

Нехай у нас є два набори даних: результати контрольної групи та експериментальної групи. Для них сформовано гіпотези: *нульова гіпотеза (H0)* – середні значення двох груп однакові; *альтернативна гіпотеза (H1)* – середні значення двох груп різні.

Потрібно порівняти середні значення цих двох груп, щоб з'ясувати, чи існують суттєві відмінності. Для порівняння середніх двох

незалежних груп **вибірок** застосовується **t-тест**. Він перевіряє гіпотезу, чи є статистично значущі відмінності між середніми цих двох груп.

```
# Створимо два набори даних (наприклад, оцінки студентів)
control_group <- c(70, 75, 80, 85, 90, 88, 79, 84)
experiment_group <- c(85, 90, 92, 91, 89, 93, 87, 94)
# Проведемо t-тест для незалежних вибірок
t_test_result <- t.test(control_group, experiment_group)
# Виведемо результати тесту
t_test_result
```

За результатами роботи функції експериментатор отримає такі дані: t статистику, стандартну помилку, p-значення (це значення допомагає визначити, чи є різниця статистично значущою) та довірчий інтервал для різниці середніх.

```
Welch Two Sample t-test
data: control_group and experiment_group
t = -3.3557, df = 9.759, p-value = 0.007544
alternative hypothesis: true
difference in means is not equal to 0
95 percent confidence interval:
-14.579403 -2.920597
sample estimates:
mean of x mean of y
81.375 90.125
```

За результатами проведених обчислень $p\text{-value} = 0.007544$, що менше за рівень значущості 0.05, отже, нульову гіпотезу (H_0) відкидають. Можна зробити висновок, що різниця між групами є статистично значущою.

У випадку, якщо порівнюється більше ніж дві групи, використовують **метод аналізу варіацій**. ANOVA дає змогу порівняти середні значення більше ніж двох груп, щоб визначити, чи є статистично значущі відмінності між ними.

Під час експерименту сформовано три групи студентів, які пройшли різні методи навчання, потрібно порівняти їхні оцінки: Формуємо гіпотези. *Нульова гіпотеза (H_0)* – усі середні значення груп однакові. *Альтернативна гіпотеза (H_1)* – хоча б одна група має інше середнє значення.

```
# Створимо три набори даних (наприклад, оцінки студентів у трьох групах)
```

```
group1 <- c(75, 78, 82, 85, 80)
group2 <- c(85, 88, 90, 92, 91)
group3 <- c(70, 72, 75, 78, 77)
# Об'єднаємо ці дані в один датафрейм для зручності
data <- data.frame(scores = c(group1, group2, group3),
  group = factor(rep(c("Group 1", "Group 2", "Group 3"), each = 5)))
# Проведемо ANOVA
anova_result <- aov(scores ~ group, data = data)
# Виведемо результати ANOVA
summary(anova_result)
```

Результати дисперсійного аналізу, проведеного ANOVA, свідчать, що $F(2,12) = 25$ та $p\text{-value} < 0.001$, це дає підстави експериментатору відхилити нульову гіпотезу (H_0).

```
Df Sum Sq Mean Sq F value Pr(>F)
group      2  558.4    279.20  25
5.25e-05 ***
Residuals 12  134.0     11.17
---
Signif. codes:  0 '***' 0.001 '**'
0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Для уточнення нагадаємо, що $p\text{-value} = 5.25e-5$ означає, що у випадку, якщо гіпотеза H_0 : істинна, ймовірність отримати значення $F = 25$ дорівнює $5.25e-5$, а це набагато менше за 0.05.

Автору дослідження потрібно визнати, що хоча б одна з груп має статистично значущі відмінності від інших. **F-статистика** вказує на величину відмінностей між групами порівняно з варіацією всередині груп. Результат ANOVA вказує на те, що є статистично значущі відмінності. Для визначення, між якими саме групами є відмінності, застосовують **пост-хок тестування** (наприклад, тест Тьюкі). Тест Тьюкі порівнює всі пари груп і вказує, чи є статистично значущі різниці між ними.

```
# Проведемо тест Тьюкі для парних порівнянь
tukey_result <- TukeyHSD(anova_result)
# Виведемо результати
tukey_result
```

Ці методи допомагають зробити об'єктивні висновки про вплив різних факторів на результати педагогічного експерименту.

Так, для проведення кореляційного аналізу можна використовувати різні методи, зокрема коефіцієнт кореляції Пірсона, Спірмена та Кендалла.

Припустимо, у нас є датафрейм `St_Performance` і ми хочемо дослідити кореляцію між змінними `Hours_Studied` (Вивчені години) та `Performance_Index` (індекс успішності)

```
data(St_Performance)
cor(St_Performance$Hours_Studied,
St_Performance$Performance_Index,
method = "pearson")
cor.test(St_Performance$Hours_Studied,
St_Performance$Performance_Index,
method = "pearson")
```

Візуалізація кореляційної матриці дасть змогу краще зрозуміти взаємозв'язки, для цього використовують пакет `corrplot`. Цей аналіз допоможе визначити, які змінні між собою найбільш пов'язані та якого характеру цей зв'язок.

```
install.packages("corrplot")
library(corrplot)
M <- cor(St_Performance)
corrplot(M, method = "circle")
```

Регресійний аналіз використовується для моделювання залежності між змінними, що дає змогу прогнозувати значення залежної змінної на основі незалежних змінних. В R є багато інструментів для проведення регресії, зокрема лінійна регресія (Linear Regression) та логістична регресія (Logistic Regression).

Лінійна регресія застосовується, коли залежна змінна є неперервною. Використовуючи лінійну регресію, моделюємо залежність індекса ефективності (`Performance_Index`) від параметрів робочого дня студента у наборі даних `St_Performance`:

```
data(St_Performance)
# Побудова моделі лінійної регресії
model <- lm(Performance_Index ~ Hours_Studied + Previous_Scores + Sleep_Hours + Papers_Practiced, data = St_Performance)
# Виведення результатів
summary(model)
```

Як результат, експериментатор отримає параметри лінійної залежності та показники, що характеризують цю модель. Для візуалізації регресійної моделі використовують такий код:

```
library(ggplot2)
ggplot(mtcars, aes(x = hp, y = mpg))
geom_point() geom_smooth(method = "lm",
col = "blue") theme_minimal()
```

Логістична регресія застосовується, коли залежна змінна є категорійною (бінарною, наприклад 0 або 1). Під час аналізу даних важливо визначити, чи є результати статистично

значущими. Для цього застосовують тестування гіпотез за допомогою різних статистичних критеріїв.

Так, для порівняння середніх значень двох груп використовується t-критерій Стьюдента (t-test). Наприклад, одновибірковий t-тест дає змогу перевірити, чи середнє значення вибірки значно відрізняється від заданого значення (наприклад, 20).

```
data <- c(22, 19, 25, 23, 20, 18, 24)
t.test(data, mu = 20)
```

Критерій χ^2 (хи-квадрат) використовується для аналізу зв'язку між категорійними змінними. Наприклад, потрібно перевірити за результатами навчання, чи є зв'язок між спеціальністю студентів та формою навчання.

```
tbl <- table(St_Performance$typ,
St_Performance$form)
chisq.test(tbl)
```

Отримані результати оцінюємо за величиною p-value.

- p-value < 0.05 → змінні залежні.
- p-value ≥ 0.05 → немає зв'язку між змінними.

Якщо дані не відповідають нормальному розподілу, використовують непараметричний тест. U-критерій Манна – Уїтні (непараметричний аналог t-тесту). Наприклад, потрібно вивчити, чи існує зв'язок між результатами навчання та формою навчання.

```
wilcox.test(Performance_Index ~ form,
data = St_Performance)
```

Для кращого аналізу та презентації даних у RStudio можна використовувати графіки, діаграми та таблиці. Найпопулярнішими інструментами в базовій версії R є пакет `ggplot2`. Так, для візуалізації розподілу змінної використовується гістограма (Histogram).

```
library(ggplot2)
ggplot(St_Performance, aes(x = Performance_Index)) +
geom_histogram(fill = "blue", bins = 10,
alpha = 0.7) + theme_minimal()
```

Іноді корисно налаштувати кількість інтервалів для гістограми, щоб точніше побачити розподіл.

```
# Побудова гістограми з визначеною
кількістю бінів
hist(x, breaks=5, main="Гістограма
з налаштованими біном", xlab="Значення",
ylab="Частота", col="lightgreen",
border="black")
```

Для побудови кумулятивної гистограми додається параметр `freq = FALSE` та використовується функція `lines()` для побудови кумулятивної лінії.

```
# Побудова кумулятивної гистограми
hist(x, freq = FALSE, main="Кумулятивна гистограма",
      xlab="Значення", ylab="Щільність", col="lightcoral",
      border="black")
lines(density(x), col="blue")
```

Після проведення статистичних тестів важливо коректно інтерпретувати результати та зробити висновки щодо гіпотез. Розглянемо, як аналізувати основні статистичні показники та застосовувати їх у педагогічній практиці.

Аналіз результатів статистичних тестів t-тест (порівняння середніх значень двох груп). Перевіримо, чи відрізняється успішність учнів у двох класах.

```
t.test(classA_scores, classB_scores)
```

За отриманими результатами показника p-value робимо висновок. У випадку $p\text{-value} < 0.05 \rightarrow$ існує статистично значуща різниця між групами, якщо $p\text{-value} > 0.05 \rightarrow$ немає значущої різниці, тобто результати схожі. Якщо $p\text{-value} < 0.05$, можна зробити висновок, що одна з навчальних методик значно впливає на успішність учнів.

Експериментатору потрібно уточнити, чи залежить успішність учнів від типу школи (міська чи сільська). Для вивчення зв'язків між категорійними змінними використовують Хі-квадрат тест ()

```
chisq.test(table(school_type,
                 success_rate))
```

У випадку, якщо $p\text{-value} < 0.05 \rightarrow$ є статистично значущий зв'язок між типом школи й успішністю. За значення $p\text{-value} > 0.05 \rightarrow$ змінні незалежні, тобто школа не впливає на рівень успішності.

Результати регресійного аналізу дають можливість досліднику зробити цілий ряд висновків. Аналізуємо, як кількість годин навчання впливає на результати тесту.

```
model <- lm(test_score ~ study_hours, data = Sts_data)
summary(model)
```

Результати дослідження дають змогу зробити ряд висновків. За результатами обчислення R-squared роблять висновок про відсоток поясненої варіації (чим ближче до 1, тим краща модель). Коефіцієнт Estimate () показує, на скільки балів зміниться успішність зі збільшенням годин навчання. Значення p-value для кожного предиктора вказує, що за $p < 0.05$ змінна значуща. Це важливо для перевірки гіпотез. Якщо study_hours має високий коефіцієнт і $p < 0.05$, можна рекомендувати збільшення навчального часу для покращення результатів.

Проведення аналізу статистичних даних педагогічного дослідження – це складний динамічний процес. На кожному циклі опрацювання даних дослідник приймає рішення про задачі наступного етапу. Середовище RStudio забезпечує достатній набір інструментів для проведення аналітичного дослідження будь якої складності.

Висновки. Цілісна методична система проведення аналізу результатів педагогічного дослідження достатньо чітко сформована в науковій літературі. Використання RStudio відкриває широкі можливості для якісного аналізу даних, візуалізації результатів та автоматизації обчислювальних процесів. Пакет надає доступ до ефективних різнопланових інструментів опрацювання статистичних даних, що сприяє підвищенню точності та надійності досліджень у сфері освіти.

Застосування RStudio дає змогу дослідникам ефективно працювати на всіх етапах процесу опрацювання результатів педагогічного експерименту: від збору та реєстрації експериментальних даних до їх інтерпретації залежно від результатів обчислень. Кожна з розглянутих функцій та бібліотек середовища RStudio містить певну кількість параметрів, що розширює їхні можливості, але вони не ввійшли в наше дослідження. Перспективу нашого дослідження вбачаємо в деталізації методики кожного з етапів педагогічного дослідження з орієнтацією на використання RStudio.

Список використаних джерел

1. Архипова С. П. Використання методів математичної статистики для перевірки результатів соціально-педагогічного експерименту. *Вісник Черкаського університету. Серія Педагогічні науки*. 2009. № 146. С. 9–14. URL: http://intellect-invest.org.ua/rus/pedagog_editions_e-magazine_pedagogical_scince_arhiv_pn_nl.
2. Гончаренко С. У. Педагогічні дослідження: методологічні поради молодим науковцям. Київ – Вінниця : ДОВ «Вінниця». 2008. 278 с.
3. Кофанов О. Є., Солнцев С. О., Зозульов О. В. Програмування із використанням R у статистичних та маркетингових дослідженнях: навчально-методичний комплекс дисципліни : навч. посіб. для студентів спеціальності 075 «Маркетинг». Київ : КПП ім. Ігоря Сікорського, 2023. 204 с. URL: <https://ela.kpi.ua/server/api/core/bitstreams/f78aa74c-7d8d-4c84-87eb-18a4aec57476/content>.
4. Роїк М. В., Присяжнюк О. І., Денисюк В. О. Огляд програмних засобів статистичного аналізу даних. *Електронний журнал «Ефективна економіка»*. 2017. № 7. URL: <http://www.economy.nayka.com.ua/?op=1&z=5676>.
5. Руденко В. М. Математико-статистичні методи в педагогічних дослідженнях : навчальний посібник. Рівне : Волинські обереги, 2012. 584 с. URL: <https://textbook.com.ua/statistika/1473453148>.
6. Чорна А. В., Осадчий В. В. Обробка експериментальних даних за результатами педагогічного експерименту засобами комп'ютерних статистичних пакетів. *Педагогічний дискурс : зб. наук. праць*. № 14. 2013. С. 453–457.
7. Tikka S., Kopra J., Heinäniemi M., López-Pernas S., Saqr M. Introductory Statistics with R for Educational Researchers. *Learning Analytics Methods and Tutorials*. Springer, Cham. 2024, pp. 121–150. https://doi.org/10.1007/978-3-031-54464-4_5.

References

1. Arkhypova, S.P. (2009). Vykorystannia metodiv matematychnoi statystyky dlia perevirky rezultativ sotsialno-pedahohichnoho eksperymentu [Using mathematical statistics methods to verify the results of a socio-pedagogical experiment]. *Visnyk Cherkaskoho universytetu. Seriya Pedahohichni nauky*, 146, 9–14. Retrieved from http://intellect-invest.org.ua/rus/pedagog_editions_e-magazine_pedagogical_scince_arhiv_pn_nl [in Ukrainian].
2. Honcharenko, S.U. (2008). Pedahohichni doslidzhennia: metodolohichni porady molodym naukovtsiam [Pedagogical Research: Methodological Advice for Young Scientists]. Kyiv – Vinnytsia: DOV “Vinnytsia”. 278 p. [in Ukrainian].
3. Kofanov, O.Ye., Solntsev, S.O., & Zozulov, O.V. (2023) Prohramuvannia iz vykorystanniam R u statystychnykh ta marketynhovykh doslidzhenniakh: navchalno-metodychnyi kompleks dystsypliny : navch. posib. dlia studentiv spetsialnosti 075 “Marketynh” [Programming Using R in Statistical and Marketing Research: Educational and Methodological Complex of the Discipline: A Textbook for Students Majoring in 075 “Marketing”]. Kyiv: KPI im. Ihoria Sikorskoho. 204 p. Retrieved from <https://ela.kpi.ua/server/api/core/bitstreams/f78aa74c-7d8d-4c84-87eb-18a4aec57476/content> [in Ukrainian].
4. Roik, M.V., Prysiazhniuk, O.I., & Denysiuk, V.O. (2017). Ohliad prohramnykh zasobiv statystychnoho analizu danykh [Overview of statistical data analysis software tools]. *Elektronnyi zhurnal “Efektyvna ekonomika”*, 7. Retrieved from <http://www.economy.nayka.com.ua/?op=1&z=5676> [in Ukrainian].
5. Rudenko, V.M. (2012). Matematyko-statystychni metody v pedahohichnykh doslidzhenniakh: navchalnyi posibnyk [Mathematical and statistical methods in pedagogical research: A Textbook]. Rivne: Volynski oberehy. 584 p. Retrieved from <https://textbook.com.ua/statistika/1473453148> [in Ukrainian].
6. Chorna, A.V., & Osadchyi, V.V. (2013). Obrobka eksperymentalnykh danykh za rezultatamy pedahohichnoho eksperymentu zasobamy kompiuternykh statystychnykh paketiv [Processing of experimental data based on the results of a pedagogical experiment using computer statistical packages]. *Pedahohichniy dyskus: zb.nauk. prats*, 14. 453–457 [in Ukrainian].
7. Tikka, S., Kopra, J., Heinäniemi, M., López-Pernas, S., Saqr, M. (2024). Introductory Statistics with R for Educational Researchers. *Learning Analytics Methods and Tutorials*. Springer, Cham., 121–150 https://doi.org/10.1007/978-3-031-54464-4_5 [in English].