TESTING THE DIFFERENCES IN AVERAGE HEIGHT BETEWEEN BOYS AND GIRLS

A PROJECT REPORT

Submitted by

NANDHINI K (2303811724322075)

in partial fulfillment of requirements for the award of the course

AGB1252 - FUNDAMENTALS OF DATASCIENCE USING R

in

ARTIFICIAL INTELLIGENCE AND DATA SCIENCE

K. RAMAKRISHNAN COLLEGE OF TECHNOLOGY

(An Autonomous Institution, affiliated to Anna University Chennai and Approved by AICTE, New Delhi)

SAMAYAPURAM – 621 112

JUNE- 2025

K. RAMAKRISHNAN COLLEGE OF TECHNOLOGY (AUTONOMOUS)

SAMAYAPURAM – 621 112

BONAFIDE CERTIFICATE

Certified that this project report on **" TESTING THE DIFFERENCES IN AVERAGE HEIGHT BETWEEN BOYS AND GIRLS"** is the bonafide work of **NANDHINI K (2303811724322075)** who carried out the project work during the academic year 2024 - 2025 under my supervision.

SIGNATURE

Dr.T. AVUDAIAPPAN, M.E., Ph.D.,

HEAD OF THE DEPARTMENT

ASSOCIATE PROFESSOR

Department of Artificial Intelligence

K. Ramakrishnan College of Technology (Autonomous)

Samayapuram–621112.

SIGNATURE

Ms.S. MURUGAVALLI., M.E., (Ph.D.).,

SUPERVISOR

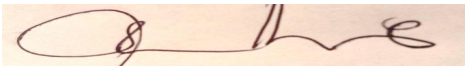
ASSISTANT PROFESSOR

Department of Artificial Intelligence

K. Ramakrishnan College of Technology (Autonomous)

Samayapuram–621112.

Submitted for the viva-voce examination held on 02.06.2025



INTERNAL EXAMINER

EXTERNAL EXAMINER

DECLARATION

I declare that the project report on **"TESTING THE DIFFERENCES IN AVERAGE HEIGHT BETWEEN BOYS AND GIRLS"** is the result of original work done by me and best of our knowledge, similar work has not been submitted to **"ANNA UNIVERSITY CHENNAI"** for the requirement of Degree of **BACHELOR OF TECHNOLOGY**. This project report is submitted on the partial fulfilment of the requirement of the completion of the course **AGB1252 – FUNDAMENTALS OF DATASCIENCE USING R.**

Signature

NANDHINI K

Place: Samayapuram

Date: 02.06.2025

ACKNOWLEDGEMENT

It is with great pride that I express our gratitude and in-debt to our institution "**K.Ramakrishnan College of Technology (Autonomous)**", for providing us with the opportunity to do this project.

I glad to credit honourable chairman **Dr. K. RAMAKRISHNAN**, **B.E.,** for having provided for the facilities during the course of our study in college.

I would like to express our sincere thanks to our beloved Executive Director **Dr. S. KUPPUSAMY, MBA, Ph.D.,** for forwarding to our project and offering adequate duration in completing our project.

I would like to thank **Dr. N. VASUDEVAN, M.Tech., Ph.D.,** Principal, who gave opportunity to frame the project the full satisfaction.

I whole heartily thanks to **Dr. T. AVUDAIAPPAN, M.E., Ph.D.,** Head of the department, **ARTIFICIAL INTELLIGENCE** for providing his encourage pursuing this project.

I wish to convey our profound and heartfelt gratitude to our esteemed project supervisor **Ms. S. MURUGAVALLI., M.E., (Ph.D.)**, Department of **ARTIFICIAL INTELLIGENCE** for her incalculable suggestions, creativity, assistance and patience, which motivated us to carry out this project.

I render our sincere thanks to Course Coordinator and other staff members for providing valuable information during the course.

I wish to express our special thanks to the officials and Lab Technicians of our departments who rendered their help during the period of the work progress.

INSTITUTE

Vision:

- To serve the society by offering top-notch technical education on par with global standards.

Mission:

- Be a center of excellence for technical education in emerging technologies by exceeding the needs of industry and society.
- Be an institute with world class research facilities.
- Be an institute nurturing talent and enhancing competency of students to transform them as all – round personalities respecting moral and ethical values.

DEPARTMENT

Vision:

- To excel in education, innovation, and research in Artificial Intelligence and Data Science to fulfil industrial demands and societal expectations.

Mission

- To educate future engineers with solid fundamentals, continually improving teaching methods using modern tools.
- To collaborate with industry and offer top-notch facilities in a conducive learning environment.
- To foster skilled engineers and ethical innovation in AI and Data Science for global recognition and impactful research.
- To tackle the societal challenge of producing capable professionals by instilling employability skills and human values.

PROGRAM EDUCATIONAL OBJECTIVES (PEO)

- **PEO1:** Compete on a global scale for a professional career in Artificial Intelligence and Data Science.
- **PEO2:** Provide industry-specific solutions for the society with effective communication and ethics.
- **PEO3** Enhance their professional skills through research and lifelong learning initiatives.

- **PROGRAM SPECIFIC OUTCOMES (PSOs)**
- **PSO1:** Capable of finding the important factors in large datasets, simplify the data, and improve predictive model accuracy.
- **PSO2:** Capable of analyzing and providing a solution to a given real-world problem by designing an effective program.

PROGRAM OUTCOMES (POs)

Engineering students will be able to:

1. **Engineering knowledge:** Apply the knowledge of mathematics, science, engineering fundamentals, and an engineering specialization to the solution of complex engineering problems.

2. **Problem analysis:** Identify, formulate, review research literature, and analyze complex engineering problems reaching substantiated conclusions using first principles of mathematics, natural sciences, and engineering sciences

3. **Design/development of solutions:** Design solutions for complex engineering problems and design system components or processes that meet the specified needs with appropriate consideration for the public health and safety, and the cultural, societal, and environmental considerations

4. **Conduct investigations of complex problems:** Use research-based knowledge and research methods including design of experiments, analysis and interpretation of data, and synthesis of the information to provide valid conclusions

5. **Modern tool usage:** Create, select, and apply appropriate techniques, resources, and modern engineering and IT tools including prediction and modeling to complex engineering activities with an understanding of the limitations

6. **The engineer and society:** Apply reasoning informed by the contextual knowledge to assess societal, health, safety, legal and cultural issues and the consequent responsibilities relevant to the professional engineering practice

7. **Environment and sustainability:** Understand the impact of the professional engineering solutions in societal and environmental contexts, and demonstrate the knowledge of, and need for sustainable development

8. **Ethics:** Apply ethical principles and commit to professional ethics and responsibilities and norms of the engineering practice.
9. **Individual and team work:** Function effectively as an individual, and as a member or leader in diverse teams, and in multidisciplinary settings.
10. **Communication:** Communicate effectively on complex engineering activities with the engineering community and with society at large, such as, being able to comprehend and write effective reports and design documentation, make effective presentations, and give and receive clear instructions.
11. **Project management and finance:** Demonstrate knowledge and understanding of the engineering and management principles and apply these to one's own work, as a member and leader in a team, to manage projects and in multidisciplinary environments.
12. **Life-long learning:** Recognize the need for, and have the preparation and ability to engage in independent and life-long learning in the broadest context of technological change.

ABSTRACT

The understanding height variations between boys and girls is essential across multiple disciplines, including healthcare, sports, and education. Identifying whether a statistically significant difference exists in their average heights provides valuable insights into growth trends, nutrition planning, and biological development. The project employs R programming for statistical testing, ensuring precise, reliable, and reproducible outcomes. By collecting and analyzing height data, the study applies hypothesis testing techniques, such as the independent t-test, to examine gender-based differences. The results can guide decisions in personalized nutrition, physical education programs, and medical diagnostics, making the research beneficial for both academic and practical applications.

ABSTRACT WITH POs AND PSOs MAPPING

CO 5 : BUILD DATABASES FOR SOLVING REAL-TIME PROBLEMS.

ABSTRACT	POs MAPPED	PSOs MAPPED
The understanding height variations between boys and girls is essential across multiple disciplines, including healthcare, sports, and education. Identifying whether a statistically significant difference exists in their average heights provides valuable insights into growth trends, nutrition planning, and biological development. The project employs R programming for statistical testing, ensuring precise, reliable, and reproducible outcomes. By collecting and analyzing height data, the study applies hypothesis testing techniques, such as the independent t-test, to examine gender-based differences. The results can guide decisions in personalized nutrition, physical education programs, and medical diagnostics, making the research beneficial for both academic and practical applications.	**PO1 -3** **PO2 -3** **PO3 -3** **PO4 -3** **PO5 -3** **PO6 -3** **PO7 -3** **PO8 -3** **PO9 -3** **PO10 -3** **PO11-3**	**PSO1 -3** **PSO2 -3**

Note: 1- Low, 2-Medium, 3- High

TABLE OF CONTENTS

CHAPTER 1

INTRODUCTION

1.1 OBJECTIVE

The primary objective of this project is to determine whether there is a statistically significant difference in the average height of boys and girls. By analyzing gender-based height data, the study aims to uncover patterns related to biological growth and nutritional factors. This information is useful for making informed decisions in fields like healthcare, sports training, and educational development. Accurate results are ensured by using statistical methods supported by R programming. The focus is on delivering reliable, data-backed conclusions. Ultimately, the project helps bridge understanding between physical growth patterns and their real-world applications.

1.2 OVERVIEW

Height is a critical measure of growth, and studying its variation between boys and girls can reveal valuable insights. Such differences may be influenced by genetic, nutritional, and environmental factors. By analyzing collected height data based on gender, the study contributes to a better understanding of developmental trends. The findings have wide applications in sports selection, education planning, and health assessments. A systematic and data-driven approach ensures precision in the analysis. This overview highlights the broader importance of studying gender-specific growth patterns.

1.3 DATA SCIENCE RELATED CONCEPTS

R programming plays a central role in this project by enabling data import, cleaning, analysis, and visualization. Statistical techniques, particularly the independent t-test, are used to test the difference in mean heights. Through exploratory data analysis (EDA), trends and variations in the dataset are identified. Key machine learning concepts like data preprocessing and feature analysis are also applied. Visual tools such as boxplots and histograms help in better interpretation of the results. While the focus is on statistics, the project sets a foundation for future use of ML in similar studies.

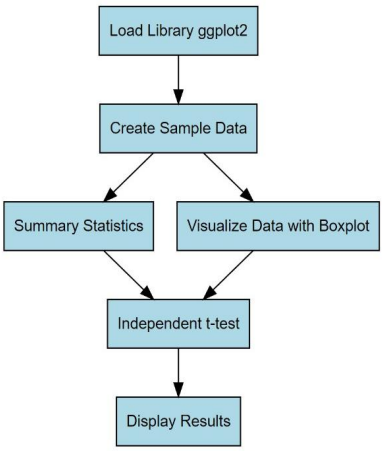
CHAPTER 2

PROJECT METHODOLOGY

2.1 PROPOSED WORK

The process involves conducting a statistical analysis to compare the average height of boys and girls. The first step is to collect a reliable dataset containing gender-wise height information. After data collection, preprocessing will be done to clean and organize the data for analysis. Descriptive statistics will be calculated to summarize the data, followed by data visualization using graphs such as boxplots and histograms for better understanding. An independent t-test will be performed using R programming to determine if the difference in average heights between boys and girls is statistically significant. The final step is to interpret the results and draw meaningful conclusions that can be applied in educational, sports, and healthcare settings.

2.2 BLOCK DIAGRAM



CHAPTER 3
MODULE DESCRIPTION

3.1 DATA COLLECTION MODULE

This module relevant data on the height of boys and girls is gathered from reliable sources such as public datasets, school records, or surveys. The dataset must include gender labels and height values to ensure proper comparison. The quality of the data is crucial, as errors at this stage can affect the accuracy of results. The collected data is stored in a structured format, like CSV or Excel, for easy use in R programming. This forms the foundation for the analysis in subsequent modules.

3.2 DATA PREPROCESSING MODULE

Once the data is collected, it undergoes preprocessing to remove inconsistencies, missing values, and outliers. This step ensures the dataset is clean and suitable for statistical analysis. Data types are checked and converted if necessary, and gender labels are standardized (e.g., converting "M"/"F" to "Boy"/"Girl"). Any irrelevant columns are removed to focus solely on gender and height. This module prepares the data for accurate and meaningful testing.

3.3 STATISTICAL TESTING MODULE

This module focuses on applying statistical methods to determine if there is a significant difference in the average heights of boys and girls. The independent t-test is used to compare the means of the two groups. Assumptions of normality and equal variance are checked before performing the test. The t-test provides a p-value that helps determine statistical significance. This result will confirm whether the height difference is due to random chance or represents a true variation.

3.4 DATA VISUALIZATION MODULE

To support the analysis and improve interpretability, data visualization techniques are applied using R programming. Graphs such as boxplots, histograms, and bar charts are created to show the distribution and comparison of heights between boys and girls. These visuals help in identifying patterns, trends, and outliers in the data. Well-designed plots make the results easy to understand for both technical and non-technical audiences. This module enhances the clarity and impact of the findings

CHAPTER 4
CONCLUSION & FUTURE SCOPE

CONCLUSION

The analysis reveals notable differences in height between boys and girls, reflecting underlying biological and developmental factors. Understanding these differences plays a crucial role in designing targeted nutritional and health interventions, as well as optimizing educational and sports programs. Employing statistical methods in R programming provided a robust and reliable framework for evaluating the data, ensuring that the conclusions drawn are both accurate and meaningful.

FUTURE SCOPE

Expanding the research to include a broader demographic range and additional variables such as socioeconomic status, diet, and physical activity could offer deeper insights into the factors influencing height differences. Applying similar statistical techniques to other growth indicators or health parameters may further enrich knowledge in pediatric growth and development. Moreover, creating user-friendly R-based applications could support practitioners in real-time monitoring and decision-making related to child growth and well-being

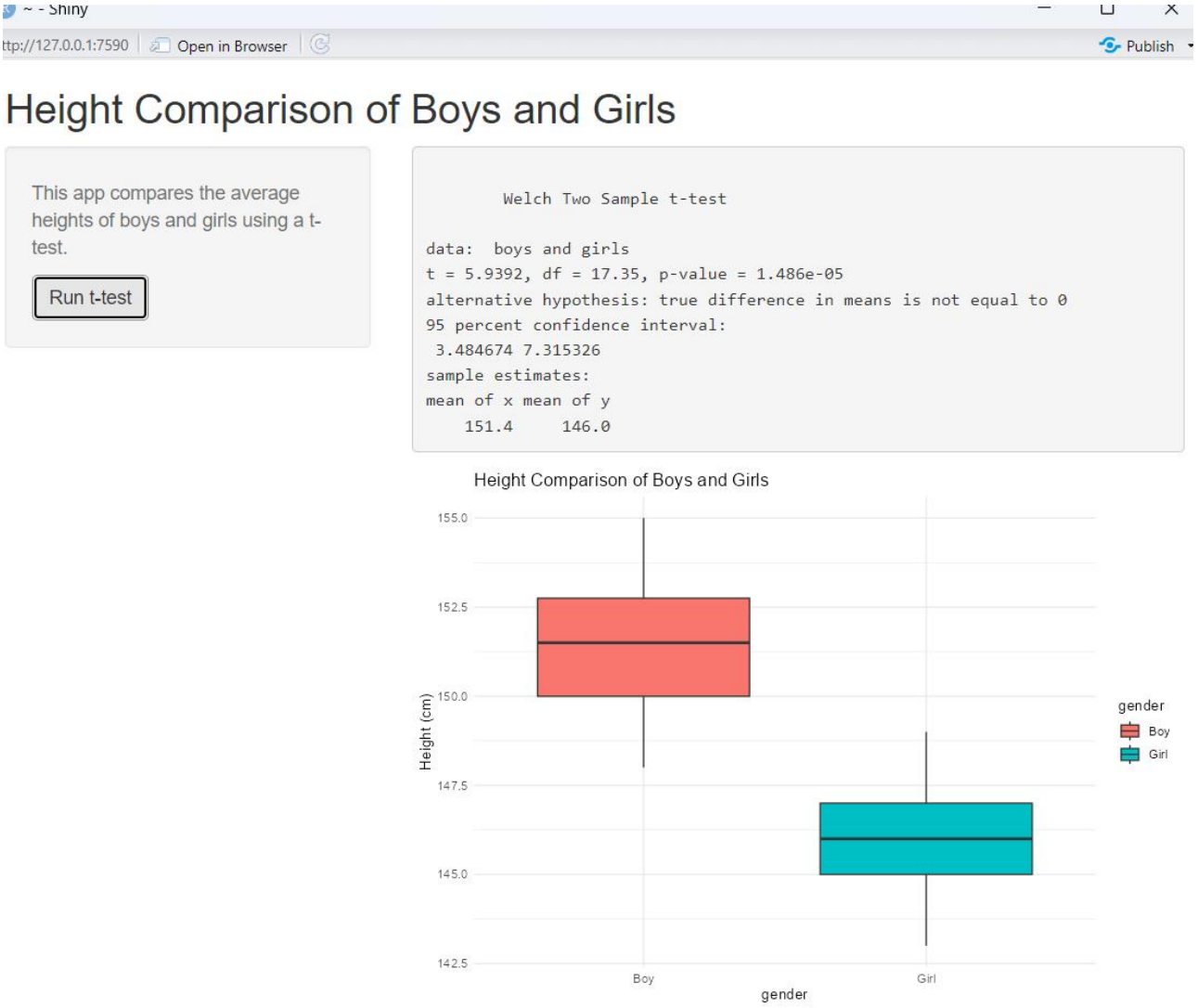
APPENDICES

APPENDIX A – SOURCE CODE

```
library(shiny)
library(tidyverse)

data <- data.frame(
  gender = c(rep("Boy", 10), rep("Girl", 10)),
  height = c(150, 152, 155, 149, 151, 153, 148, 154, 150, 152,
             145, 147, 149, 144, 146, 148, 143, 147, 145, 146)
)
ui <- fluidPage(
  titlePanel("Height Comparison of Boys and Girls"),

  sidebarLayout(
    sidebarPanel(
      helpText("This app compares the average heights of boys and girls using a t-test."),
      actionButton("runTest", "Run t-test")
    ),
     mainPanel(
      verbatimTextOutput("testResult"),
      plotOutput("boxPlot")
    )
  )
)
server <- function(input, output) {

  boys <- filter(data, gender == "Boy")$height
  girls <- filter(data, gender == "Girl")$height

  observeEvent(input$runTest, {
    output$testResult <- renderPrint({
      t.test(boys, girls)
    })
  })
```

```
  output$boxPlot <- renderPlot({
    ggplot(data, aes(x = gender, y = height, fill = gender)) +
      geom_boxplot() +
      labs(title = "Height Comparison of Boys and Girls", y = "Height (cm)") +
      theme_minimal()
  })
}
shinyApp(ui = ui, server = server)
```

APPENDIX B – SCREENSHOTS



REFERENCES:

1. **Crawley, M. J. (2013).** ***Statistics****: An Introduction Using R* (2nd ed.). Wiley.
2. **Garn, S. M., & Clark, D. C. (1976).** Sex differences in growth patterns of children and adolescents. *American Journal of Physical Anthropology*, 44(2), 215–222. https://doi.org/10.1002/ajpa.1330440213
3. **WHO Multicentre Growth Reference Study Group. (2006):** *WHO Child Growth Standards: Methods and Development*. World Health Organization.