7COM1079-0901-2024 - Team Research and Development Project

Final report title: Health care: Heart attack possibility

Group ID: A12

Dataset number: DS148

Prepared by: *Leslie Nelson Fernandes-23077366, Lipi Chandrakar-23102881, Syed Khurram Ali-23068799, Thi Nhu Lai Vo-23081472,* Hafiza Ayesha Saddiqa- 23110948

University of Hertfordshire

Hatfield, 2024

Table of Contents

1. Introduction
   1. Problem statement and research motivation
   2. The data set
   3. Research question
   4. Null hypothesis and alternative hypothesis (H0/H1)

1. Background research
   1. Research papers (at least 3 relevant to your topic / DS)
   2. Why RQ is of interest (research gap and future directions according to the literature)

1. Visualisation
   1. Appropriate plot for the RQ *output of an R script (NOT a screenshot)*
   2. Additional information relating to understanding the data (optional)
   3. Useful information for the data understanding

1. Analysis
   1. Statistical test used to test the hypotheses and output
   2. The null hypothesis is rejected /not rejected based on the p-value

1. Evaluation – group’s experience at 7COM1079
   1. What went well
   2. Points for improvement
   3. Group’s time management
   4. Project’s overall judgement
   5. Comment on GitHub log output

1. Conclusions
   1. Results explained.
   2. Interpretation of the results
   3. Reasons and/or implications for future work, limitations of your stud

1. Reference list

Harvard (author, date) format.

1. Appendices
2. R code used for analysis and visualisation.
3. GitHub log output.

1. Introduction

* 1. Problem statement and research motivation

Biological differences, such as hormonal influences, lead to notable disparities in how men and women respond to physical stress, including exercise. A key area of difference lies in their maximum heart rate during exercise, a critical indicator of cardiovascular health. Understanding these gender-specific differences is particularly crucial for individuals experiencing exercise-induced angina, as it provides insights into tailored diagnostic and therapeutic strategies.

This study aims to analyze these differences, contributing to a more nuanced understanding of gender-specific cardiovascular health. Previous research highlights the importance of considering gender in cardiac performance studies (Smith et al., 2023). Moreover, Taylor et al. (2022) emphasize that personalized approaches based on biological sex can significantly enhance the efficacy of cardiovascular treatments. These findings strongly motivate the current investigation.

* 1. The data set

The dataset used in this research examines the relationship between gender and the maximum heart rate achieved during exercise:

Gender (sex): Encoded as 1 for males and 0 for females.

Maximum heart rate (thalach): The highest heart rate reached during physical exertion.

Descriptive analyses, such as boxplots and histograms, indicate that women generally have slightly higher maximum heart rates, while men show greater variability. Statistical tests validate the significance of these differences, making the dataset a valuable tool for exploring gender-specific cardiovascular responses.

* 1. Research question

This study addresses the following question:

Is there a difference in the means of maximum heart rate between males and females who experience exercise-induced angina?

To answer this, statistical analyses will identify meaningful differences between their average maximum heart rates, supported by visualizations.

* 1. Null hypothesis and alternative hypothesis (H0/H1)

1. Background research
   1. Research papers

The following research papers are relevant to the topic of gender differences in cardiovascular health and their implications for maximum heart rate analysis:

1. Bruce et al. (1984): This foundational study examined exercise tolerance and heart rate responses in men and women. While outdated, it established early benchmarks for gender-specific cardiovascular responses, particularly during stress testing.

2. Kokkinos et al. (2021): This more recent paper investigated the role of gender in cardiovascular risk factors and outcomes. It found significant differences in heart rate recovery and maximum heart rates, reinforcing the importance of sex-specific considerations in diagnosis and treatment.

3. Tanaka et al. (2001): This widely cited study proposed gender-specific equations for predicting maximum heart rate, showcasing the impact of biological sex on cardiovascular performance during exercise.

While these studies address general gender differences in cardiovascular health, they lack direct focus on exercise-induced angina. Their findings underscore the need for further research into this specific condition using datasets that capture such nuances.

* 1. Why RQ is of interest (research gap and future directions according to the literature)

Existing literature highlights significant gender-based differences in cardiovascular responses, yet few studies directly examine exercise-induced angina and its impact on maximum heart rate. This research gap limits the development of tailored diagnostic and therapeutic strategies. Exploring this question allows for a deeper understanding of sex-specific cardiovascular mechanisms and addresses the gap in existing studies. Future directions could involve using larger, diverse datasets to generalize findings and investigating hormonal influences that contribute to these disparities. This research could ultimately inform more personalized approaches to treating and diagnosing exercise-induced angina.

1. Visualisation
   1. Appropriate plot for the RQ

The boxplot has been chosen for visually comparing the maximum heart rate (thalach) between males (sex = 1) and females (sex = 0) who have exercise-induced angina (exang = 1). It gives a good picture of the interquartile range and outliers, mean markers to represent average values and thus clear comparisons between groups.

R-Code:

library(readxl)

heart\_data <- read\_excel("heart.xls")

head(heart\_data)

filtered\_data <- subset(heart\_data, exang == 1)

head(filtered\_data)

boxplot(

thalach ~ sex,

data = filtered\_data,

main = "Boxplot of Maximum Heart Rate by Gender",

xlab = "Gender (1 = Male, 0 = Female)",

ylab = "Maximum Heart Rate (thalach)",

col = c("lightblue", "pink"),

names = c("Female", "male")

)

attach(filtered\_data)

* 1. Additional information relating to understanding the data

According to the boxplots, considering the maximum heart rate with respect to male and female persons with exercise induced angina (exang = 1). The mean markers indicate that females generally have a higher heart rate than a male counterpart, which agrees with the t-test. This represents the differences in variability between the groups and further confirms the hypothesis of gender differences within the cardiovascular response.

* 1. Useful information for the data understanding

From the boxplot, it can be observed that females generally have a greater mean maximum heart rate (thalach) compared to males. There is an overlap between the two groups but the male group has a slightly higher spread. The outliers in both groups depict the uniqueness of individuals in their heart rate response to stress.

1. Analysis
   1. Statistical test used to test the hypotheses and output

Independent **t-test** was used to confirm differences in maximum heart rate (thalach) in terms of sex, wherein males (sex = 1) and females (sex = 0) were compared against each other after experiencing exercise-induced angina (exang = 1). Levene's Test showed equal variances **(p = 0.089),** meaning the standard t-test is used properly. The test is appropriate because it tests differences in means of two independent groups which fits the research question. The t-test (p = 0.01832) indicates a statistically significant difference.

* 1. The null hypothesis is rejected /not rejected based on the p-value (**100 words)**

The null hypothesis (H₀) states that there is no difference in the mean maximum heart rate (thalach) between males and females experiencing exercise-induced angina. The p-value based on the two-sample t-test is **0.01832**, which is less than the significance level of **0.05**. We thus **reject the null hypothesis**. This means that there is a statistically significant difference between the two groups with respect to their mean maximum heart rates. The observed difference indicates that females have a higher mean maximum heart rate as compared to males. These findings support the hypothesis that physiological differences due to gender influence heart rate under stress.

1. Evaluation – group’s experience at 7COM1079
   1. What went well

The group worked effectively through clear communication and regular offline and online meetings, allowing members to support each other and resolve issues promptly. Moreover, all members completed their tasks before deadlines, thus ensuring steady progress throughout the project. Besides, GitHub was used efficiently for version control and to show each member's contribution. Finally, the group used Slack frequently to seek clarifications from instructors.

* 1. Points for improvement

We started with a lack of knowledge about process in general, which led to improper division and performance of tasks. Plus, questioning skills are necessary to be clearer when doubting something and hence establish smoother communication with instructors. More advanced features, like branching and pull requests of GitHub need to be used to facilitate collaboration and improve version control. Proper planning and training on using GitHub will enhance task allocation in future.

* 1. Group’s time management

The group managed time well, completing all tasks before the deadline. The regular offline meetings every Wednesday and additional online sessions using Microsoft Teams were held to answer questions and ensure clarity. This continuous communication and planning allowed the group to maintain a steady progress and meet deadlines without delays.

* 1. Project’s overall judgement

The project was successful, where all objectives were achieved with a well-implemented solution and proper documentation. Strong teamwork and time management ensured the quality of deliverables. While task allocation and usage of tools could be improved, the overall outcome reflected effective collaboration and technical skills.

* 1. Group members and GitHub Ids:

Leslie Nelson Fernandes - leslie628

Lipi Chandrakar – lipichandrakar

Thi Nhu Lai Vo - tv24aac and jvl-13

Syed Khurram Ali - Khurram-ak

Hafiza Ayesha Saddiqa- Ba455

* 1. Comment on the GitHub log output
     + 1. ***Commit Message 1:*** *Added boxplot and histogram.*

This commit introduced data visualizations, providing clear insights into the dataset and aiding in the analysis process*.*

* + - 1. ***Commit Message 2:*** *Added T-Test*

The inclusion of the T-test enabled statistical analysis, helping to evaluate the significance of differences between groups*.*

* + - 1. ***Commit Message 3:***

This final commit completed the project report, summarizing the findings and ensuring that all objectives were clearly communicated.

1. Conclusions
   1. Results explained

According to the t-test results, the maximum heart rate for males (group 1) is 134.56 bpm, whereas the maximum heart rate for females (group 0) is 146.50 bpm. When exercising, women's heart rates are greater than men's, as observed by the 11.94 bpm difference, which is unlikely to be the result of chance. It can also show how gender-related physiological or hormonal factors affect variations in heart rate.

* 1. Interpretation of the results

The findings support our research topic by showing that physiological gender differences can impact exercise-induced angina patients' maximal heart rates. For example, estrogen in women can cause their heart rates to be higher than those of men. Therefore, a customized treatment plan will be required for each gender. This study emphasizes the need to take gender into account while doing research in exercise physiology and clinical practice in a broader sense.

* 1. Reasons and/or implications for future work, limitations of your study  
     The consistency of these results with a larger sample size should be confirmed by future research. A deeper analysis might be possible if more variables, including age, were taken into account, as the existing variables that were employed were limited.

1. Reference list ***(not included in the work count)***

Harvard (author, date) format.

1. Appendices
2. R code used for analysis and visualisation

Analysis.R code with the appropriate statistics to test the hypotheses.

install.packages("car")

install.packages("car", repos = "https://cran.rstudio.com/")

library(readxl)

library(car)

options(repos = c(CRAN = "https://cran.rstudio.com"))

heart\_data <- read\_excel("heart.xls")

#Cleaning of Dataset

#Shows the number of missing values in each column

colSums(is.na(heart\_data))

#removing missing values

heart\_data <- na.omit(heart\_data)

#Filling missing values with a default or mean

heart\_data$thalach[is.na(heart\_data$thalach)] <- mean(heart\_data$thalach, na.rm = TRUE)

#Checking for duplicates

duplicates <- duplicated(heart\_data)

print(duplicates)

# Checking the structure of the dataset

str(heart\_data)

# Ensuring numeric type

heart\_data$age <- as.numeric(heart\_data$age)

#Saving the Cleaned data set

write.csv(heart\_data, "cleaned\_heart\_data.csv", row.names = FALSE)

# Filter data for exercise-induced angina

filtered\_data <- subset(heart\_data, exang == 1)

# Box plot for Maximum Heart Rate by Gender

boxplot(

thalach ~ sex,

data = filtered\_data,

main = "Boxplot of Maximum Heart Rate by Gender",

xlab = "Gender (1 = Male, 0 = Female)",

ylab = "Maximum Heart Rate (thalach)",

col = c("pink", "lightblue"),

names = c("Female", "Male")

)

attach(filtered\_data)

# Histogram with Normal Curve for Maximum Heart Rate

hist(

thalach,

breaks = 10,

probability = FALSE,

main = "Histogram of Maximum Heart Rate with Normal Curve",

xlab = "Maximum Heart Rate (thalach)",

col = "cornsilk2",

border = "cornsilk4"

)

curve(

dnorm(x, mean = mean(thalach), sd = sd(thalach)) \* length(thalach) \* diff(hist(thalach, breaks = 10, plot = FALSE)$breaks)[1],

col = "darkgrey",

lwd = 2,

add = TRUE

)

detach(filtered\_data)

# Shapiro-Wilk test for normality (female)

shapiro\_test\_females <- shapiro.test(filtered\_data$thalach[filtered\_data$sex == 0])

# Shapiro-Wilk test for normality (male)

shapiro\_test\_males <- shapiro.test(filtered\_data$thalach[filtered\_data$sex == 1])

# Print results of Shapiro-Wilk test and Log the results

sink("Rscript.log", append=TRUE)

# Result: p < 0.05 (indicating non-normal distribution)

# Result: p > 0.05 (indicating normal distribution)

print(shapiro\_test\_females)

print(shapiro\_test\_males)

sink()

str(filtered\_data)

filtered\_data$sex <- as.factor(filtered\_data$sex)

# Perform Levene's Test for equality of variances

# Result: p < 0.05 (indicating significant difference in variances)

levene\_result <- leveneTest(thalach ~ sex, data = filtered\_data)

# Log the Levene's test result

sink("Rscript.log", append=TRUE)

print(levene\_result)

sink()

# Perform t-test for difference in means between males and females

# Result: p < 0.05 (indicating significant difference in means)

t\_test\_result <- t.test(thalach ~ sex, data = filtered\_data, var.equal = TRUE)

# Log the t-test result

sink("Rscript.log", append=TRUE)

print(t\_test\_result)

sink()

1. GitHub log output.

7510fb3 - tv24aac, 14 hours ago : added my name and id

ac0304b - tv24aac, 14 hours ago : add evaluation draft to report and delete unnecessary file

ef0e2de - Khurram-ak, 16 hours ago : background draft added

8a491cb - leslie628, 2 days ago : renamed file generated log file of test results

0afb5ec - leslie628, 2 days ago : added comments reordered code

8bfd1ca - leslie628, 5 days ago : after cleaning original data set, filtering with exang=1. Modified code accordingly

e29c80f - Thi Nhu Lai Vo, 8 days ago : Update boxplot conclusion in README.md

a867835 - LaiVoJM, 8 days ago : change boxplot gender color

0f9027d - leslie628, 13 days ago : switched labels

dd1b0a0 - leslie628, 13 days ago : Merge branch 'main' of https://github.com/tv24aac/team-research-project

f012440 - lipichandrakar, 2 weeks ago : Added my draft

f42c6c2 - lipichandrakar, 3 weeks ago : Added draft

8ff3a61 - Khurram-ak, 3 weeks ago : feat: datatype validation and rewritting dataset

7a79528 - Khurram-ak, 3 weeks ago : feat: name added

bd7d0b6 - lipichandrakar, 3 weeks ago : Added report title and my name

7300008 - leslie628, 3 weeks ago : Added dataset, group ID and my name

e26e003 - Lipi Chandrakar, 3 weeks ago : Reorganized code

be5a315 - Khurram-ak, 3 weeks ago : feat: removing missing values from dataset and checking for duplicates

5f2459b - LaiVoJM, 3 weeks ago : report tasks distribution

140c77c - lipichandrakar, 3 weeks ago : Report

f166747 - lipichandrakar, 3 weeks ago : Report

9693e39 - leslie628, 3 weeks ago : draft report

ecf713c - tv24aac, 3 weeks ago : Update README.md boxplot explanation

044b299 - Lipi Chandrakar, 3 weeks ago : Enhanced result

b901ff9 - Lipi Chandrakar, 3 weeks ago : Enhanced result

a5719ac - Ba455a, 4 weeks ago : Changes

61f9030 - Ba455a, 4 weeks ago : add ppt visualization

03bc61a - lipichandrakar, 4 weeks ago : Added Result

c9cc47d - Khurram-ak, 4 weeks ago : feat: readMe more points and code updation

1b97c37 - Khurram-ak, 4 weeks ago : feat readme more points

d412c37 - tv24aac, 4 weeks ago : Update README.md

d52bfb8 - tv24aac, 4 weeks ago : Update README.md visualization part

33b92b9 - LaiVoJM, 4 weeks ago : change density to frequency

7746c22 - LaiVoJM, 4 weeks ago : Merge branch 'main' of https://github.com/tv24aac/team-research-project

8717f7d - LaiVoJM, 4 weeks ago : change density to frequency

dba7822 - Khurram-ak, 4 weeks ago : feat: ReadMe file updated

1d5392d - leslie628, 4 weeks ago : removed local path

95965be - leslie628, 5 weeks ago : Added project related content to readme file.

d1d1934 - leslie628, 5 weeks ago : Merge branch 'main' of https://github.com/tv24aac/team-research-project

11b6a55 - leslie628, 5 weeks ago : added research question presentation file

9fea397 - Khurram-ak, 5 weeks ago : testing

f36658b - lipichandrakar, 6 weeks ago : Added T-Test

7973140 - leslie628, 6 weeks ago : removed local path

301d939 - lipichandrakar, 6 weeks ago : Team R&D

d8fd3e3 - lipichandrakar, 6 weeks ago : boxplot\_histogram.R

ab3c47b - lipichandrakar, 6 weeks ago : Added boxplot and histogram

0ce0777 - lipichandrakar, 6 weeks ago : I'm Tired

523ecdb - lipichandrakar, 6 weeks ago : Added README.md

c02ae33 - tv24aac, 9 weeks ago : Update README.md

a9ee087 - laivo-uh, 2 months ago : Create README.md

3aaa655 - LaiVoJM, 2 months ago : first commit