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, [Name and ID of other group members]*

***Please make sure*** *the document spelled correctly (including image labels, section headings, and table of contents). Please use correct punctuation.*

*Make sure your report is grammatically correct.*

University of Hertfordshire

Hatfield, 2024

Table of Contents *add page numbers here*

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.

*The list below outlines the chapter/subchapter numbers, names, word count limits, and explanations of what to write in each section.*

1. Introduction

* 1. Problem statement and research motivation **(100 words)**
* *What is the problem in the area we want to learn more about (motivation for study).*
* *Use at least one citation from related literature for top marks.*
  1. The data set **(75 words)**
  2. Research question **(50 words).** *Explain how you are going to answer your RQ.* **(50 words)**
  3. Null hypothesis and alternative hypothesis (H0/H1) **(100 words)**

1. Background research
   1. Research papers (at least 3 relevant to your topic / DS) **(200 words)**

* *Was the data set used for some research papers?* *Reference at least 3 relevant research papers to your topic / DS.*
  1. Why RQ is of interest (research gap and future directions according to the literature) **(100 word**s)

1. Visualisation
   1. Appropriate plot for the RQ *output of an R script (NOT a screenshot)* (**50 words)**

* *Explain the choice of the plot.*
* *Anything on the plot from R is not counted towards word count limit*
* *Make sure that the plot is from output of an R script (NOT a screenshot)*
* *Make sure that the plot has a caption or title, X and Y-axis labels, legend if appropriate and units.*
* *Make sure the title or caption and axis labels are informative.*
  1. Additional information relating to understanding the data (optional) (**50 words)**
* *Per plot: explain the purpose and insights.*
  1. Useful information for the data understanding (**50 words)**
* *Summarise key observations from the plot.*

1. Analysis
   1. Statistical test used to test the hypotheses and output (**75 words)**

* *Explain the choice of the test.*
* *Make sure the test is appropriate for the RQ and data.*
  1. The null hypothesis is rejected /not rejected based on the p-value (**100 words)** *(interpret the results)*

1. Evaluation – group’s experience at 7COM1079
   1. What went well **(75 words)**
   2. Points for improvement **(75 words)**
   3. Group’s time management (**50 words)**
   4. Project’s overall judgement (**50 words)**
   5. Note any changes to group since submission of Assignment 1. Add new or amended GitHub Ids for new members **(75 words, write only if applies to your group arrangements)**
   6. Comment on the GitHub log output **(50 words)**

*Please comment on the GitHub log output, and refer to it as being placed into**Appendix B.*

*From your Git log, select the three most significant commits during this project and include the following for each:*

1. ***Commit Message:*** *[Insert Commit Message] Brief explanation of the broader impact of the change*
2. ***Commit Message:*** *[Insert Commit Message] Brief explanation of the broader impact of the change*
3. ***Commit Message:*** *[Insert Commit Message] Brief explanation of the broader impact of the change*

1. Conclusions
   1. Results explained (**75 words)**
   2. Interpretation of the results (**75 words)**

* *Interpretation of what the results mean in terms of your RQ and the effect this may have on your population and the wider context of your topic.*
  1. Reasons and/or implications for future work, limitations of your study (**50 words)**

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

Harvard (author, date) format.

1. Appendices
2. R code used for analysis and visualisation ***(not included in the word count)***

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

* ***No word count****, but ensure the code is without redundant lines, well-commented and produces the correct output.*
* *Make sure it runs (look in Rscript.log for output from a statistical test)*
* *It should compute appropriate statistics to test the hypotheses*

1. GitHub log output.

Draft-

* **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("Male", "Female")

)

attach(filtered\_data)

1. Additional information:

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 males generally have a higher heart rate than a female 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 males generally have a greater mean maximum heart rate (thalach) compared to females. 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.

* **Analysis**

1. Statistical test:

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 result:

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 males have a higher mean maximum heart rate as compared to females. These findings support the hypothesis that physiological differences due to gender influence heart rate under stress.