# Exploring the Impact of Lifestyle Factors on the Development and Progression of Heart Disease

ALBAYRAK, Melike Gülşen*

**Abstract**

This project aims to study the impact of lifestyle factors, including smoking, physical activity, diet and alcohol drinking, on the onset and progression of heart disease. The study be based on the UCI Heart Disease Dataset, which contains comprehensive clinical and demographic information on patients with suspected heart disease. Data analysis and statistical methods will be used to explore the associations between lifestyle habits and heart disease outcomes in the parts of "Data" and " Methods and Data Analysis". The project seeks to provide valuable insights into how these daily factors that influence the risk and pathway of heart diseases. The findings will contribute to the existing body of knowledge on cardiovascular health and inform the development of targeted interventions . By better understanding the role of lifestyle factors in heart disease, this research aims to improve cardiovascular health outcomes and reduce the overall heart diseases at a population level.

## 1 Introduction

Heart disease remains a significant public health concern globally, accounting for a substantial burden of morbidity and mortality. Lifestyle factors, such as smoking, physical activity, diet, and alcohol consumption, have long been recognized as key modifiable risk factors that can significantly influence the development and progression of heart disease. Understanding the impact of these lifestyle factors on cardiovascular health is crucial for effective prevention and management strategies. The purpose of this project is to investigate the interplay between lifestyle factors and the presence or severity of heart diseases, using the Heart Disease Dataset from the UCI Machine Learning Repository which is a website that i found from my searchings about the most helpful dataset providers to find suitable datasets for researchs. After my investigations for finding a good dataset, I chose the Heart Disease Dataset that has 303 observations with a set of 14 clinical and demographic variables including categorical (sex, chest pain type (cp), fasting blood sugar $> 120$ mg/dl (fbs), resting electrocardiographic results (restecg), exercise induced angina (exang), slope of the peak exercise ST segment (slope), number of major vessels colored by fluoroscopy (ca), thalassemia

*21080562, Github Repo

(thal), target), discrete numeric (maximum heart rate achieved (thalach)) and continous numeric (age, resting blood pressure (trestbps, in mm Hg), serum cholesterol (chol, in mg/dl), ST depression induced by exercise relative to rest (oldpeak)) variables. After chosing the dataset, i pursued on looking for articles related to heart diseases and i saw that how specific lifestyle behaviours contribute to the risk and progression of cardiovascular heart problems. So at last, i designed my question as "What is the impact of lifestyle factors (such as smoking, physical activity, diet, and alcohol consumption) on the development and progression of heart disease?". Then, i just focused more specifically on the articles related to my research question for the project.

### 1.1 Literature Review

In this section, discuss the articles you have read on the subject by giving references. This is a narrative citation Chang & Serletis (2013). This one is a parenthetical citation (Chang & Serletis, 2013). **Do not summarize each article individually under a separate title.** In the literature review section, **at least six** articles must be cited (Newbold et al., 2003; Verzani, 2014; Wickham, 2014; Wooldridge, 2015).

## 2 Data

In this section, discuss the source of the data set you use in your study, if you have done any operation on the raw data, these operations and the summary statistics about the data set. In this section, it is mandatory to have a table (Table 1) containing summary statistics (mean, standard deviation, minimum, maximum, etc. values) of all variables. Make the necessary references to your tables as shown in the previous sentence (Perkins et al., 1991).

R codes for the analysis should start in this section. In this section, you should include the codes that imports the data set into R and the codes that generate summary statistics.

```
library(tidyverse)
library(here)
survey <- read_csv(here("data/heartdatas.csv"))
```

Note that code options are edited in some of the code chunks in the Rmd file.

With the `echo=FALSE` option, prevent the codes from appearing in the derived pdf file and report your results in tables.

## 3 Methods and Data Analysis

In this section describe the methods that you use to achieve the purpose of the study. You should use the appropriate analysis methods (such as hypothesis tests and correlation

Table 1: Summary Statistics

| | Mean | Std.Dev | Min | Median | Max |
|---|---|---|---|---|---|
| age | 54.54 | 9.05 | 29.00 | 56.00 | 77.00 |
| ca | 0.68 | 0.94 | 0.00 | 0.00 | 3.00 |
| chol | 247.35 | 52.00 | 126.00 | 243.00 | 564.00 |
| cp | 3.16 | 0.96 | 1.00 | 3.00 | 4.00 |
| exang | 0.33 | 0.47 | 0.00 | 0.00 | 1.00 |
| fbs | 0.14 | 0.35 | 0.00 | 0.00 | 1.00 |
| oldpeak | 8.80 | 11.07 | 0.00 | 4.00 | 62.00 |
| restecg | 1.00 | 0.99 | 0.00 | 1.00 | 2.00 |
| sex | 0.68 | 0.47 | 0.00 | 1.00 | 1.00 |
| target | 0.95 | 1.23 | 0.00 | 0.00 | 4.00 |
| thal | 4.73 | 1.94 | 3.00 | 3.00 | 7.00 |
| thalach | 149.60 | 22.94 | 71.00 | 153.00 | 202.00 |
| trestbps | 131.69 | 17.76 | 94.00 | 130.00 | 200.00 |

analysis) that we covered in the class. If you want, you can also use other methods that we haven't covered. If you think some method is more suitable for the purpose of the analysis and the data set, you can use that method (Newbold et al., 2003; Verzani, 2014; Wickham, 2014; Wooldridge, 2015).

For example, if you are performing regression analysis, discuss your predicted equation in this section. Write your equations and mathematical expressions using $LaTeX$.

This section should also include different tables and plots. You can add histograms, scatter plots (such as Figure **??**), box plots, etc. Make the necessary references to your figures as shown in the previous sentence.

# 4 Conclusion

Summarize the results of your analysis in this section. Discuss to what extent your results responded to the research question you identified at the beginning and how this work could be improved in the future.

# 5 References


Chang, D., & Serletis, A. (2013). The demand for gasoline: Evidence from household survey data. *Journal of Applied Econometrics*, *29*(2), 291–313.

Newbold, P., Carlson, W. L., & Thorne, B. (2003). *Statistics for business and economics*. Pearson College Division.

Perkins, K. A., Sexton, J. E., Solberg-Kassel, R. D., & Epstein, L. H. (1991). Effects of nicotine on perceived exertion during low-intensity activity. *Medicine & Science in Sports & Exercise*.

Verzani, J. (2014). *Using R for introductory statistics*. CRC Press.

Wickham, H. (2014). *Advanced R*. CRC Press.

Wooldridge, J. M. (2015). Control function methods in applied econometrics. *Journal of Human Resources*, *50*(2), 420–445.