MFE II PROJECT

HEART DISEASE ANALYSIS

Group members:
Anusha singh (220641)
Vedansh kumawat (220489)
Shivanshi goel(220457)

HEART DISEASE ANALYSIS

Our heart is one of the most important organs in our body. Malfunctions of the heart is called heart disease or cardiac disease. Heart disease is a leading cause of death. The factors that cannot be controlled are gender, age, family history, heart shape. The controllable risk factors are high blood pressure, cholesterol level, obesity, smoking, and diabetes. Heart disease is a leading cause of death.

AGENDA

- Objective
- preprocessing of data
- data collected
- descriptive statistics and visualtisation
- hypothesis testing
- conclusion
- references

OBJECTIVE:

☑ Association between gender and risk of a heart attack:

- A Chi-square test of association was conducted to determine if there is an association between gender and the risk of a heart attack.
- The results of the test would indicate whether there is evidence to support the hypothesis that gender is associated with the likelihood of a heart attack.

☑ Mean age of males and females with a high chance of a heart attack:

- Separate analyses were performed for males and females to calculate summary statistics, such as mean and standard deviation, for their ages.
- The aim was to compare the mean ages of males and females who have a high chance of a heart attack.
- Two-sample t-tests were used to assess whether there is a significant difference in the mean ages of males and females.

Overall, the investigation aimed to provide insights into the likelihood of a heart attack based on gender and age. It involved statistical tests, summary statistics, and visualizations to support the analysis.

DATA COLLECTED

The dataset contains observations of various individuals from the United States of America, Cleveland.
The two major considerations are :

- Age: Age of individual in years.
- Sex: Gender of the individual. 0 for Female and 1 for Male. Sex variable has been renamed to Gender.

The dataset contains 76 variables from different countries such as Hungary and Switzerland but have been subset for easier analysis and the Cleveland database is the one that has been selected for this task.

PREPROCESSING OF DATA

THE CODE PROVIDED DEMONSTRATES HOW TO IMPORT A CSV FILE, RENAME A VARIABLE, CONVERT IT TO A FACTOR, GROUP THE DATA BY GENDER, AND CALCULATE SUMMARY STATISTICS USING THE SUMMARISE() FUNCTION FROM THE DPLYR PACKAGE IN R.

TO EXECUTE THIS CODE AND TO SEE THE SUMMARIZED STATISTICS, WE COPIED AND PASTED IT INTO AN R ENVIRONMENT (E.G., RSTUDIO) AFTER ENSURING THAT WE HAVE INSTALLED THE NECESSARY PACKAGE (E.G., DPLYR).

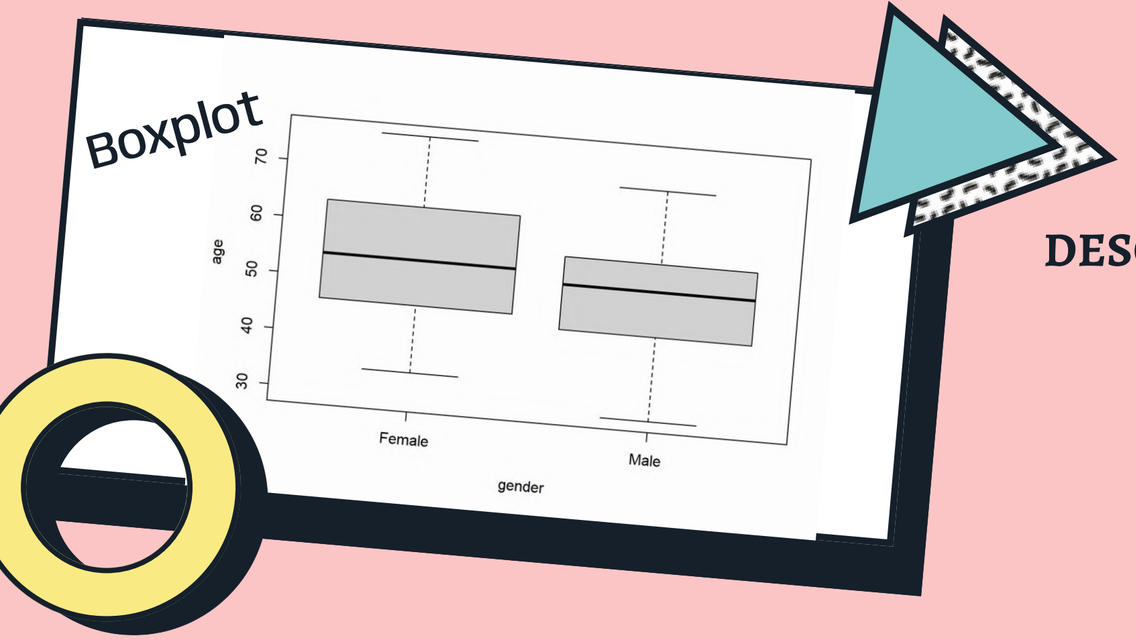
```
ble [303 x 14] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
age      : num [1:303] 63 37 41 56 57 57 56 44 52 57 ...
gender   : Factor w/ 2 levels "Female","Male": 2 2 1 2 1 2 1 2 2 2 ...
cp       : num [1:303] 3 2 1 1 0 0 1 1 2 2 ...
trestbps: num [1:303] 145 130 130 120 120 140 140 120 172 150 ...
chol     : num [1:303] 233 250 204 236 354 192 294 263 199 168 ...
fbs      : num [1:303] 1 0 0 0 0 0 0 0 1 0 ...
restecg  : num [1:303] 0 1 0 1 1 1 0 1 1 1 ...
thalach  : num [1:303] 150 187 172 178 163 148 153 173 162 174 ...
exang    : num [1:303] 0 0 0 0 1 0 0 0 0 0 ...
oldpeak  : num [1:303] 2.3 3.5 1.4 0.8 0.6 0.4 1.3 0 0.5 1.6 ...
slope    : num [1:303] 0 0 2 2 2 1 1 2 2 2 ...
ca       : num [1:303] 0 0 0 0 0 0 0 0 0 0 ...
thal     : num [1:303] 1 2 2 2 2 1 2 3 3 2 ...
target   : Factor w/ 2 levels "0","1": 2 2 2 2 2 2 2 2 2 2 ...
attr(*, "spec")=
. cols(
.   age = col_double(),
.   sex = col_double(),
.   cp = col_double(),
.   trestbps = col_double(),
.   chol = col_double(),
.   fbs = col_double(),
.   restecg = col_double(),
.   thalach = col_double(),
.   exang = col_double(),
.   oldpeak = col_double(),
.   slope = col_double(),
.   ca = col_double(),
.   thal = col_double(),
.   target = col_double()
```

DESCRIPTIVE STATISTICS AND VISUALISATION

- OUTLIERS: THERE ARE NO OUTLIERS IN THE AGE DATA FOR EITHER MALES OR FEMALES, BASED ON THE BOXPLOT VISUALIZATION.
- INTERQUARTILE RANGE (IQR): THE IQR FOR FEMALES IS GREATER THAN THAT OF MALES, INDICATING A HIGHER VARIABILITY IN AGE AMONG FEMALES.
- MEAN AGE: THE MEAN AGE FOR FEMALES IS 54.56, WHILE THE MEAN AGE FOR MALES IS 50.90. THIS SUGGESTS THAT, ON AVERAGE, FEMALES TEND TO BE OLDER THAN MALES IN THIS DATASET.
- STANDARD DEVIATION: THE STANDARD DEVIATION OF AGE FOR FEMALES IS 10.27, WHICH IS HIGHER THAN THE STANDARD DEVIATION OF AGE FOR MALES, WHICH IS 8.68. THIS FURTHER SUPPORTS THE OBSERVATION THAT THERE IS GREATER VARIABILITY IN AGE AMONG FEMALES.

THESE SUMMARY STATISTICS PROVIDE INSIGHTS INTO THE DISTRIBUTION OF AGE FOR MALES AND FEMALES IN THE DATASET, HIGHLIGHTING DIFFERENCES IN THE CENTRAL TENDENCY AND VARIABILITY BETWEEN THE GENDERS.

DESCRIPTIVE STATISTICS AND VISUALISATION



>The boxplot shows that there are no outliers in the data that requires to be dealt with.
>The interquartile range of the Female is greater than the Male which indicates there is a greater variability in the age.

> Based on the summary statistics, the Female have a higher mean age of 54.56 compared to Male at 50.90. In addition, the standard deviation of Female is 10.27, which is higher than that of Male at 8.68.

>The heart dataset is filter for individuals who have a high likelihood of heart attack.

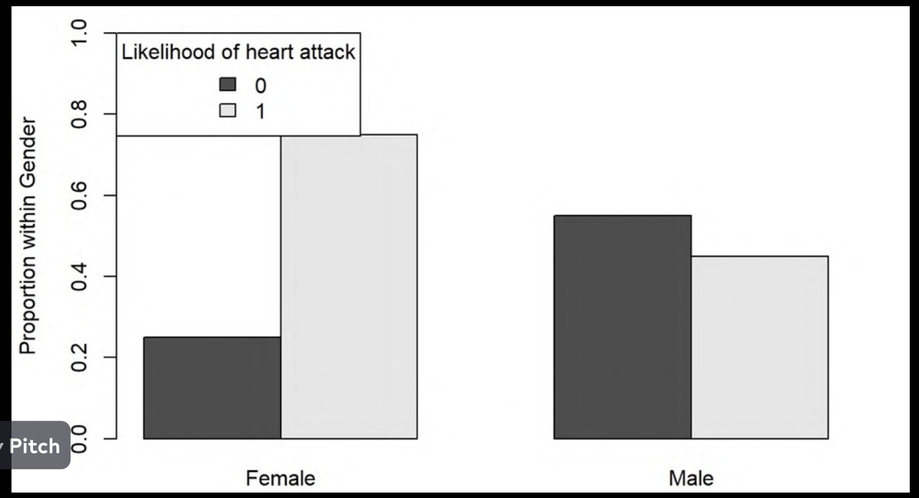
> there are greater number of observations of ages for the Male compared to the Female.

HYPOTHESIS TESTING

The following hypothesis have been set out and tested through this investigation:

>There is association between gender and risk of heart attack of an individual.

> The age of male and female who have a high chance of a heart attack are the same.



1. Chi- Square testing

> The chi-square test of association for the gender and likelihood of heart attack gives a p-value = 1.877e-06 which is less than the 0.001.

>Therefore, the null hypothesis, H0 is rejected and the chi-square test of association is statistically significant.
> The results tell us that there is no evidence of an association between the likelihood of a heart attack and the gender of the individual. Therefore, the likelihood of a heart attack is dependent of the gender of the individual.

2. Two sample t test

>The next test used will be to understand is there statistical difference in the age of Male and Female of which have a higher chance of heart attack.
>The heart dataset is filtered for individuals who have a high likelihood of heart attack.

3. Student T Test

>Found a statistically significant difference in the mean age of male and female who have high chance of heart attack. Since p value = 0.014 < 0.05, we reject the H0 and 95% CI of the estimated population difference [6.57,0.74] does not captures the H0: M1 - M2=0.
>Therefore we can conclude that the age of heart disease for male is significantly different from female.

CONCLUSION

- **In summary, the analysis consisted of two investigations:**

> The first investigation explored the association between gender and the likelihood of a heart attack. The visual representation indicated that females may be at a higher risk.

>However, the chi-square test of association showed that gender and the likelihood of a heart attack are not related. Therefore, there is no evidence to suggest that a person's gender is associated with their likelihood of experiencing a heart attack.

- **In conclusion, based on the initial analysis, there is no association between gender and the likelihood of a heart attack. However, the analysis comparing mean ages suggests a potential difference, but it is important to consider the confidence level when interpreting the results.**

REFERENCES

- WORLD HEALTH ORGANIZATION 2017, *CARDIOVASCULAR DISEASE (CVDS)* WEBPAGE (HTML FORMAT), WORLD HEALTH ORGANIZATION, MELBOURNE, HTTPS://WWW.WHO.INT/EN/NEWS-ROOM/FACT-SHEETS/DETAIL/CARDIOVASCULAR HTTPS://WWW.WHO.INT/EN/NEWS-ROOM/FACT-SHEETS/DETAIL/CARDIOVASCULAR-DISEASES-(CVDS)DISEASES-(CVDS)

- HTTPS://WWW.WHO.INT/NEWS-ROOM/FACT-SHEETS/DETAIL/THE-TOP-10-CAUSES-OF HTTPS://WWW.WHO.INT/NEWS-ROOM/FACT-SHEETS/DETAIL/THE-TOP-10-CAUSES-OF-DEATHDEATH

- HTTPS://WWW.CDC.GOV/HEARTDISEASE/FACTS.HTM

- HTTPS://NEWS.MIT.EDU/2019/MACHINE-LEARNING-SHOWS-NO-DIFFERENCE-ANGINA HTTPS://NEWS.MIT.EDU/2019/MACHINE-LEARNING-SHOWS-NO-DIFFERENCE-ANGINA-SYMPTOMS-BETWEEN-MEN-AND-WOMEN-1106 SYMPTOMS-BETWEEN-MEN-AND-WOMEN-1106

THANK YOU!