

# Statistical Analysis of Heart Failure Clinical Data Set With MATLAB

## Completed by:

Sufien Tout - 217668583

Mohamad Asif Khan - 216498453

Muhammad Ali Malik - 216225658

## Introduction:

For this project, three datasets were given to students to perform statistical analysis on using MATLAB. This group decided to choose '**Heart Failure Clinical Records Data Set**', which contains medical records of 299 patients, all of whom suffered from heart failure. The dataset comprises thirteen clinical features and can be found [here](#).

Upon inspecting the dataset, the group became curious and was determined to find out answers to the countless questions that popped into their head. The group wanted to find out whether there was a link among all the thirteen clinical features, which led them to formulate questions that would require data analysis and eventually the formulation of a business application.

This report will answer the following questions using summaries of data, graphs and statistical analysis.

1. Do women live longer than men? If so, what clinical factors may contribute to this?
2. Do men tend to get heart failures earlier in life than women?
3. Out of the total number of patients who got heart failure, what percent had diabetes, what percent were smoking? Were most of the patients young or old? What percent of the people died? What does this data show?

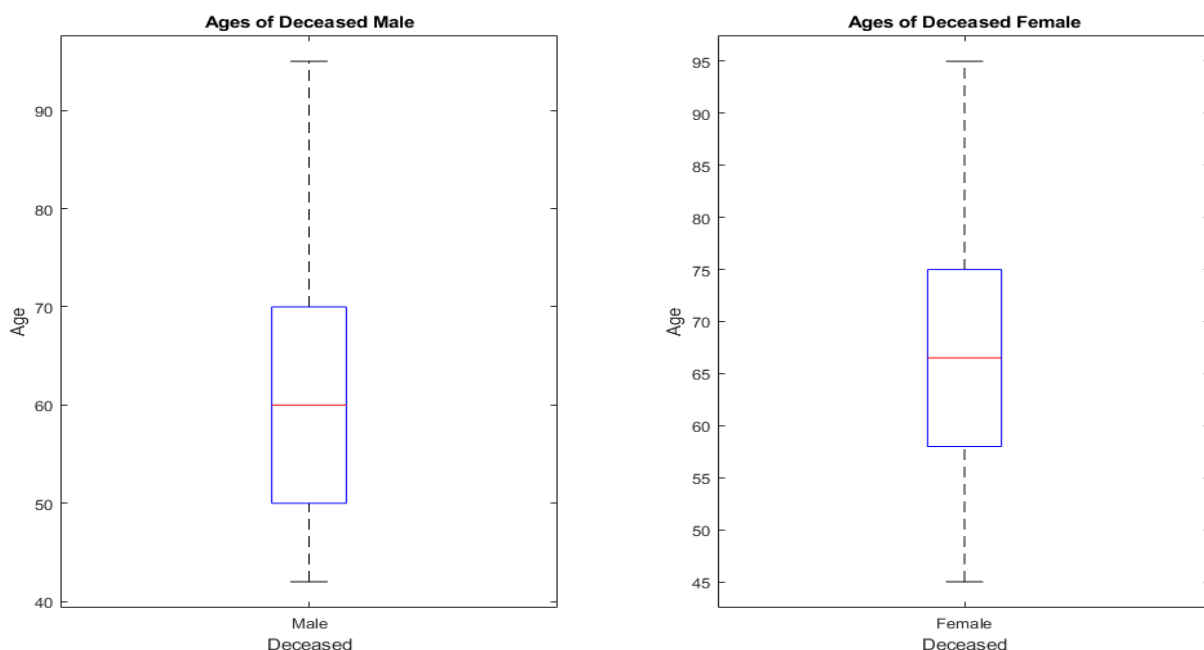
## Do women live longer than men? If so, what clinical factors may contribute to this?

It is hypothesized that women live longer than men. According to the WHO, women generally live longer than men - on average by six to eight years. The following will explore whether the hypothesis holds true by analysing the given data set and evaluating the clinical factors that affect the life expectancy of both men and women.

When analysing life expectancy, age is the major factor. Therefore, data was filtered by gender and death events, after which the ages were calculated based upon those conditions.

Conceptually, to compare the ages of deceased males and females in the dataset, boxplots can be used to statistically visualize and analyse which gender lives longer.

**Figure 1** below visualizes in form of two boxplots comparing the ages of the deceased males and females.



**Figure 1: Ages of deceased men and women**

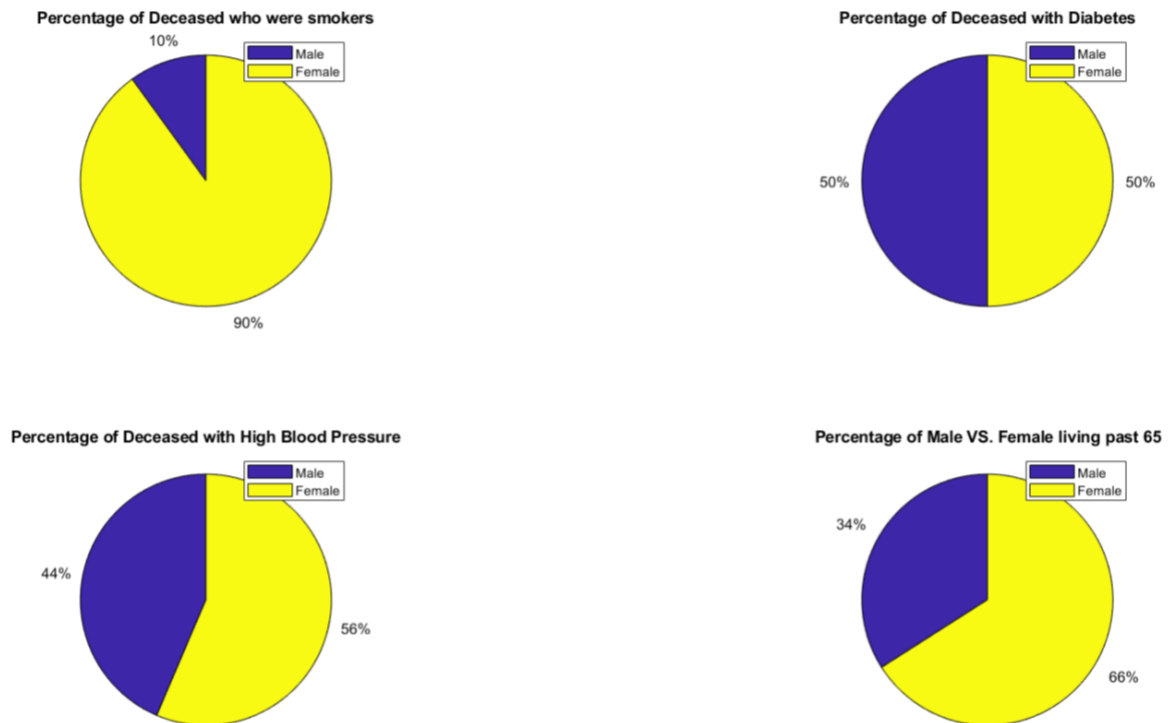
Upon inspection, the boxplot motivates one to conclude that the women died at an older age than the men, in the data set. To confirm whether this is the case, a summary table of statistical analysis can be created to take a more detailed look into the analysis.

#### Summary - Ages of Deceased - Male VS Female

SUMMARY STATISTICS(Ages of Deceased)	MALE (without outliers)	FEMALE(without outliers)
SIZE	34	62
MIN	42	45
MAX	95	95
MEAN	62.1765	66.8817
MEDIAN	60	66.5
VAR	151.9679	181.7744
STD	12.3275	13.4824

In this data set, there were 105 males and 194 females, 96 of whom deceased - 34 males and 62 females. Upon inspection, it can be observed that the average age of the deceased females is approximately 5 years more than that of their male counterparts. This agrees with the statement made by the WHO. Looking at the minimum, we can see that the youngest deceased male is still older than the youngest deceased male at 45 years of age. The female sample has resulted in a higher variance and a higher standard deviation, indicating that the ages are more spread out. For male, the median has a slightly lower value than the mean and the same can be said about the female sample too. This makes the data positively skewed for both male and female, following the Pearson Mode Skewness. It is no surprise that the 1st quartile for male is 50 and for female, it is 58. Then the 3rd quartile for female is 75 and for male, it is 70. From this data, we can conclude that it is indeed the case that women live longer than men.

Consequently, we wanted to explore the factors that might have an effect on the life-expectancy of women and men. Three factors were taken into consideration that have been shown to cause premature death: Smoking, Diabetes and High Blood Pressure.



From the data shown in the above pie charts, it can be said that there is a significantly higher percentage of deceased female smokers at 90% compared to only 10% males. As for diabetes, both genders have an equal ratio at 50%. 56% percent of the deceased women had high blood pressure with only 44% of the males having high blood pressure. It is proven that high blood pressure leads to cardiovascular disease and according to WHO, there are lower rates of cardiovascular disease in the female demographic, but the data above suggests otherwise. Despite this, as mentioned in the previous paragraph, the deceased females were older in age than the deceased males and as presented in the last pie chart, there is a higher percentage of females living past 65 compared to males. According to the CDC, smoking causes premature death and life expectancy of smokers is 10 years shorter than non-smokers but, in the data, again, it does not seem to be the case. Perhaps

there are other factors not included in the data set that are resulting in the greater life-expectancy for women.

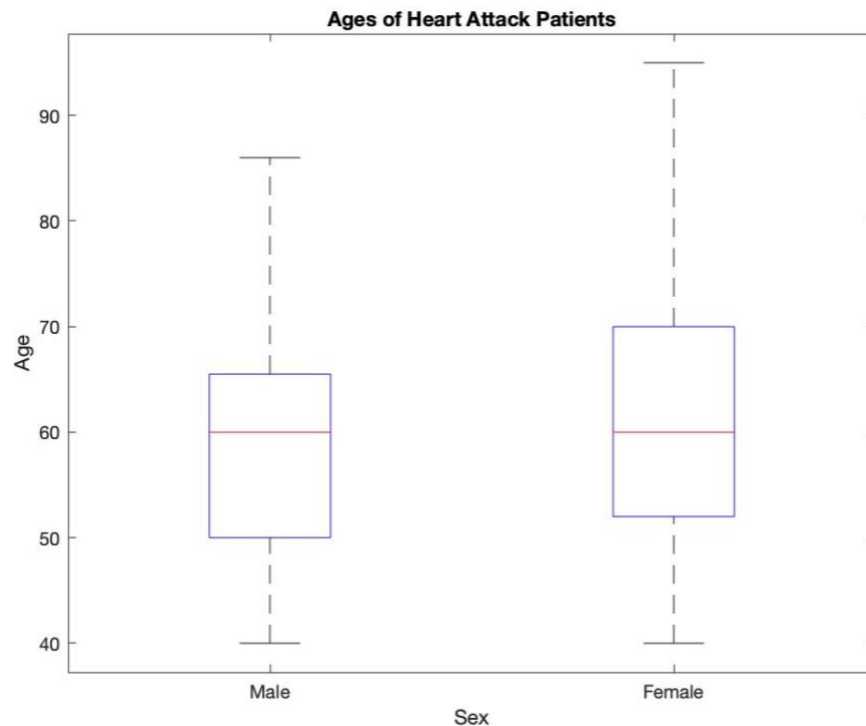
To conclude, it can be said with confidence that it is the case that women live longer than men. However, in the data set, the factors that would result in this do not seem to favour the female sample.

### **Do men tend to get heart failures earlier in life than women?**

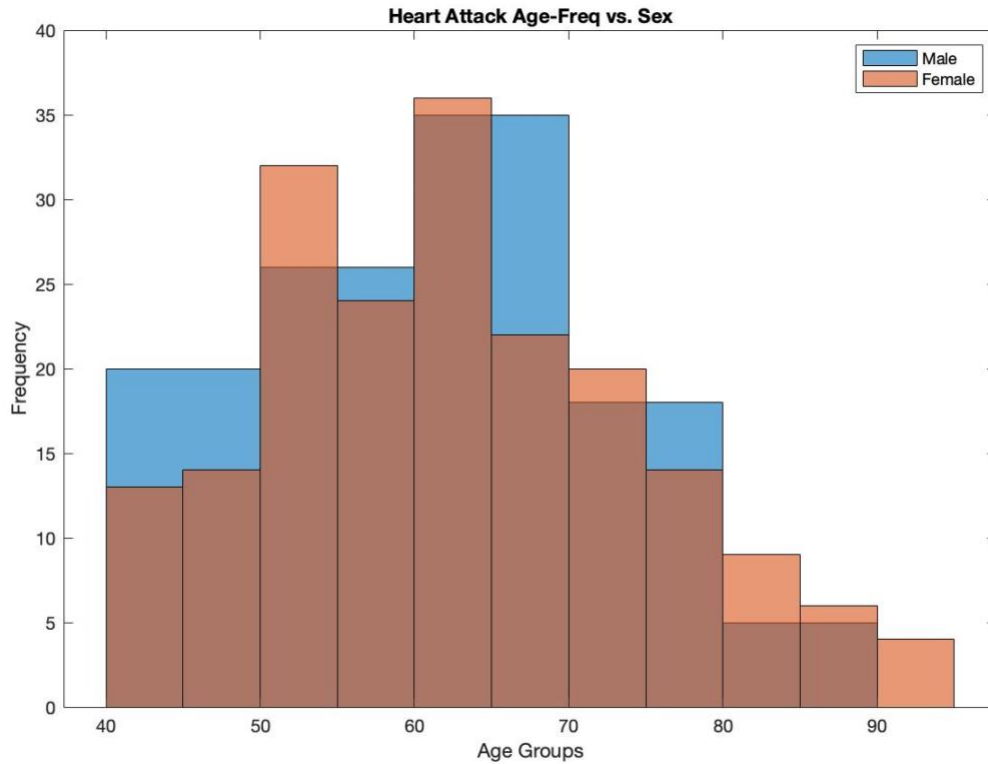
Based on a Harvard Health article, younger men are at a greater risk of heart disease than women. Heart Disease in this case is an overarching medical term that includes heart failure. With that in mind, we wanted to explore the relationship between age and sex to determine if the hypothesis applies to the patients in this dataset.

Since sex is indicated as a Boolean value in the dataset, we chose to graph the data with a histogram and boxplot. For the histogram, two series of data were constructed at various age brackets. The distribution of each series can be easily seen, where each one is slightly skewed. As for the boxplot, it gives us a quick summary of descriptive statistical information and is a sure way to indicate outliers.

SUMMARY STATISTICS	MALE	FEMALE
SIZE	104	194
MIN	40	40
MAX	90	95
MEAN	59.4391	61.4055
MEDIAN	60	60
VAR	115.4245	149.463
STD	10.7436	12.224

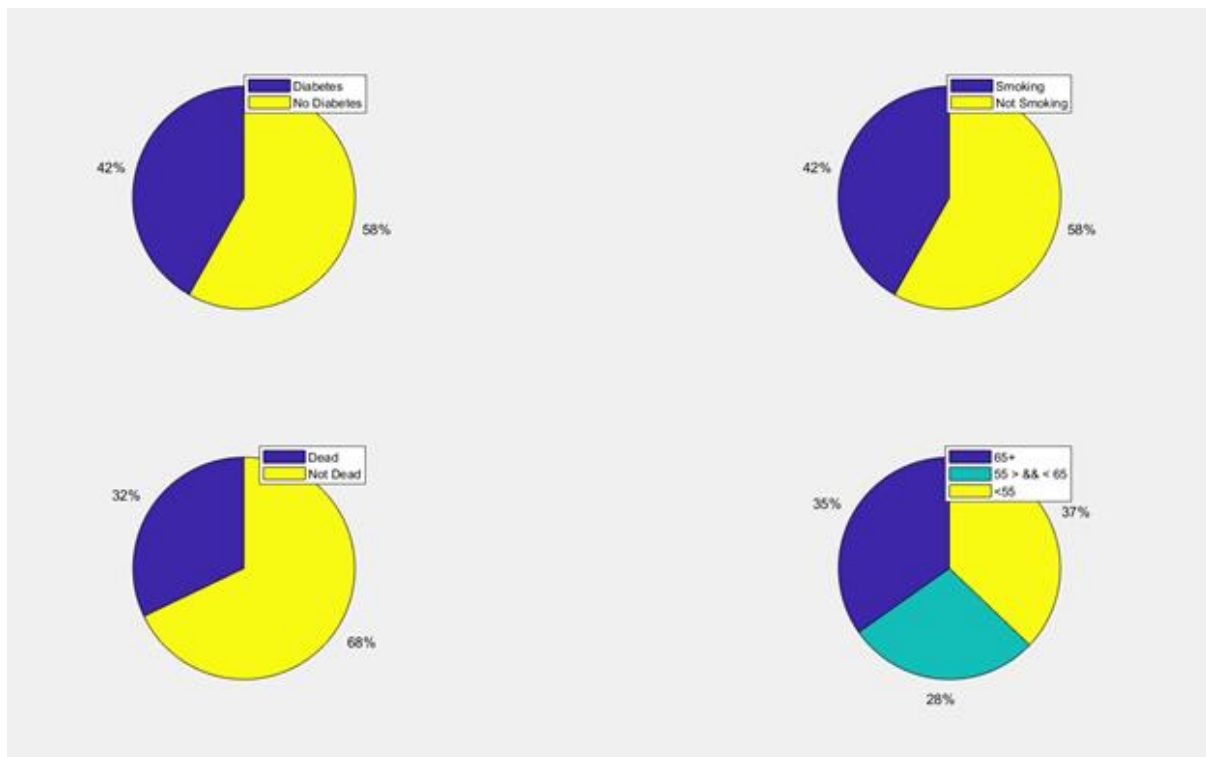


We hypothesized that men tend to get heart failure earlier in life than their female counterparts. The number of observed patients with heart failure is 299, with 194 women and 104 men excluding one outlier. Both sexes have a similar range, with men having a slightly smaller range due to the outlier. With a greater variance and standard deviation, the female sample is more spread out. Utilizing the Pearson Mode skewness, we can show that male ages are negatively skewed while the female ages are positively skewed. As for the kurtosis, males have a slightly larger value than females, yet both are less than three. This indicates that both series have a platykurtic distribution. Looking at the boxplot we can see that the 1<sup>st</sup> quartile is the age of 50 for men and 52 for women, with a quarter of the sample in a lower age for males. Both sexes have medians at the age of 60. Finally, the 3<sup>rd</sup> quartile is the age of 65 for men and 70 for women. With that in mind, these three quartiles are lower for men than women, indicating that more men tend to get heart failures earlier in life than women.



After the analysis shown above, we can conclude that men tend to get heart failures earlier in life than women. Therefore, the hypothesis does seem to apply to this dataset.

**Out of the total number of patients who got heart failure, what percent had diabetes, what percent were smoking? Were most of the patients young or old? What percent of the people died? What does this data show?**



In our project, we chose heart failure data of 299 patients that was recorded by students of Government College University, Faisalabad, Pakistan. We analysed this data in many ways, one of them was to find out how prior health problems like diabetes, habits like smoking contribute to heart failure. We also analysed how many people died because of heart failure and the number of adults and elders in the list.

As we can see from our first graph, a significant number of people around 42% had diabetes before they got heart failure. Another 58% did not have diabetes. 42% is 125 of the 299 people, this shows us that diabetes can play a great role in causing heart failure. The other 58% or 174 people who didn't have diabetes, this may be due to different reasons. We can conclude from this data that diabetes is one of the



major causes of heart failure. We must try hard by all means to avoid food and habits that cause diabetes like sugars and more.

In our second pie chart, we have the data about those who were smoking and those that don't smoke. The data proves that smoking can contribute to causing heart failure as a major number of the people who got heart failure were smokers.

Smoking is not a good habit by any means, it can cause many problems like kidney failure, different kinds of cancer and heart failure. It is said that smoking one cigarette can make your life shorter by one day.

Another important data type is the number of people who died due to heart failure. This number shows a big difference from what we taught it maybe. Heart failure kills millions of people around the world, I was expecting that most of the people who got heart failure would die of this disease, but the number shows different. It shows that only 32% heart failure patients died and 68% survived. This may be the people who got this problem for the first time, it may not show the number of people in the long term. I think most of the people on the long term will eventually die due to this problem.

Final chart shows the age of all the patients, looking at the age shows that most of the people in the data are older than 50 years. Only 37% of patients are less than 55 years old, 28% are between 55 and 65 and other 35% are older than 65 years. We can conclude from the data that heart failure is almost non-existent in youngsters, but as we age, our risk grows higher, it is very high after age 65.

## References

"Tobacco-Related Mortality." *Centers for Disease Control and Prevention*, Centers for Disease Control and Prevention, 28 Apr. 2020, [www.cdc.gov/tobacco/data\\_statistics/fact\\_sheets/health\\_effects/tobacco\\_related\\_mortality/index.htm](http://www.cdc.gov/tobacco/data_statistics/fact_sheets/health_effects/tobacco_related_mortality/index.htm). - Accessed on **2020-08-03**

"Female Life Expectancy." *World Health Organization*, WHO, 13 Dec. 2010, [www.who.int/gho/women\\_and\\_health/mortality/situation\\_trends\\_life\\_expectancy/en/](http://www.who.int/gho/women_and_health/mortality/situation_trends_life_expectancy/en/). - Accessed on **2020-08-09**

**Harvard Health Publishing. "The Heart Attack Gender Gap." Harvard Health, Apr. 2016, [www.health.harvard.edu/heart-health/the-heart-attack-gender-gap](http://www.health.harvard.edu/heart-health/the-heart-attack-gender-gap).**

Task	Finished By
Task 1: identify/handle missing/inconsistent data values	Muhammad Ali Malik Mohamad Asif Khan Sufien Tout
Task 2: Identify and remove outliers from the data	Muhammad Ali Malik Mohamad Asif Khan Sufien Tout
Task 3: Summary Table <ul style="list-style-type: none"><li>• Size</li><li>• Minimum</li><li>• 1st Quartile</li><li>• 3rd Quartile</li><li>• Mean</li><li>• Median</li><li>• Variance</li><li>• Standard Deviation</li></ul>	Muhammad Ali Malik Sufien Tout
Task 4: Plot Graphs/box-plot/hist/pie	Muhammad Ali Malik Mohamad Asif Khan Sufien Tout