
Heart Disease Research

Tanay Murdia

Columbia University

Github Repo: Heart Disease Research

Introduction

According to the data released by WHO regarding the causes of deaths worldwide from 2000-2019, heart disease has remained the leading cause of death for the last 20 years¹. The deaths by heart disease constitute about 16% of total deaths from all causes. From the data provided by Centre of Disease Control and Prevention, about 655,000 Americans die from cardiovascular disease each year, depicting the significance of statistically analyzing this disease².

Method of Data Collection

The dataset was created through a compilation of data provided by NHANES and information obtained by medical examination done by Svetlana Ulanova³. The provided dataset was quantitatively arranged as all the categorical values like cholesterol level, had been converted to numerical variables. Since there is usage of various hypothesis test, we are assuming in this data-set that all the information collected was randomized and independent of each other. One of the biggest difficulties of this research was data acquisition. The statistical analysis of this subject required a dataset that would be large enough, for the results to be applicable to a larger database.

¹ "WHO Reveals Leading Causes of Death and Disability Worldwide: 2000-2019." World Health Organization, World Health Organization, www.who.int/news/item/09-12-2020-who-reveals-leading-causes-of-death-and-disability-worldwide-2000-2019.

² Underlying Cause of Death, 1999-2019 Request." Centers for Disease Control and Prevention, Centers for Disease Control and Prevention, wonder.cdc.gov/ucd-icd10.html

³ Ulanova, S. (2019, January 20). Cardiovascular Disease dataset. Retrieved December 14, 2020, from <https://www.kaggle.com/sulianova/cardiovascular-disease-dataset>

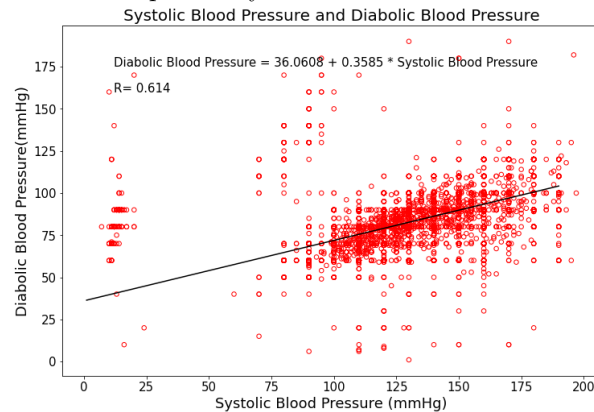
Research Question

Since cardiovascular disease constitutes 16% of total deaths from all causes around the world, this depicts the need for the analysis of the disease. This research revolves around the factors that lead to cardiovascular disease in America, where around 655,000 people die due to this disease every year. The dataset being used consists of 70,000 patient's information, which depicts that the conclusion gained from this dataset can be applied to a larger population.

Analysis

Relationship between Systolic and Diabolic Blood Pressure:

Figure 1: Scatterplot of Systolic and Diabolic Blood Pressure



Simple linear regression analysis was conducted to examine the relationship between systolic and diastolic blood pressure. A significant regression equation was found, the equation was diabolic BP = 36.06 +(0.358* systolic BP). Thus, for 1 mmHg increase in systolic blood pressure, the diabolic blood pressure increases 0.358 mmHg. The correlation coefficient between the two variables was 0.64. Therefore, there is a positive correlation between systolic and diastolic blood pressure. Through the simple linear regression, it is can be concluded that there is a medium- strong positive linear relationship between systolic and diabolic blood pressure. From the figure above, it can be observed that it is a medium-strong correlation as there are various outliers decreasing the correlation coefficient

Systolic Blood Pressure and Cardiovascular Disease:

Figure 2: Grouped Histogram of Systolic Blood Pressure with and without cardiovascular disease

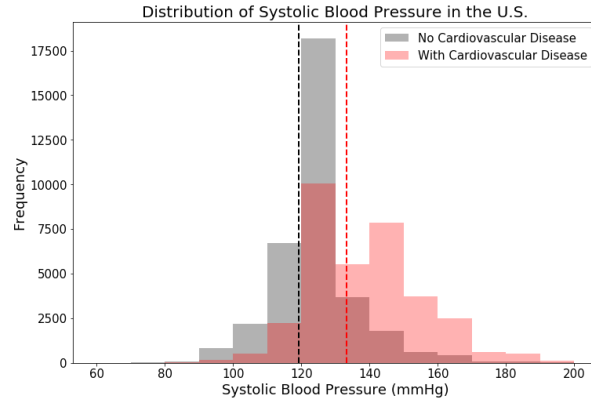
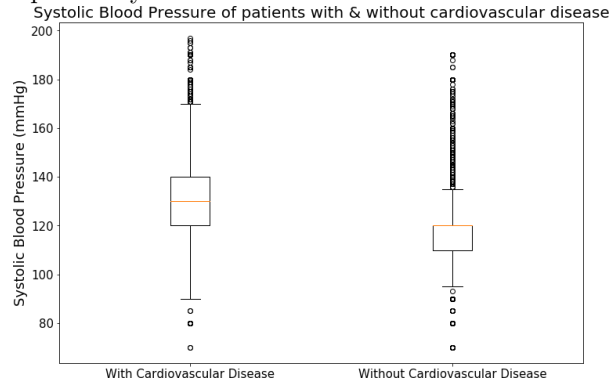


Figure 3: Grouped Boxplot of Systolic Blood Pressure with and without cardiovascular disease



The average systolic blood pressure of Americans with cardiovascular disease was 133.54 mmHg and the average systolic blood pressure of Americans without cardiovascular disease was 119.50 mmHg, which causes the average difference to be 14.04 mmHg.

This analysis demonstrates that average systolic blood pressure of a patient with cardiovascular disease is higher than average systolic blood pressure of a patient without cardiovascular disease. This can also be observed in the graph 2 and 3, where the mean and median of systolic blood pressure is higher in patients with cardiovascular disease than patients without cardiovascular disease.

Diabolic Blood Pressure and Cardiovascular Disease:

Figure 4: Grouped Histogram of Diabolic Blood Pressure with and without cardiovascular disease

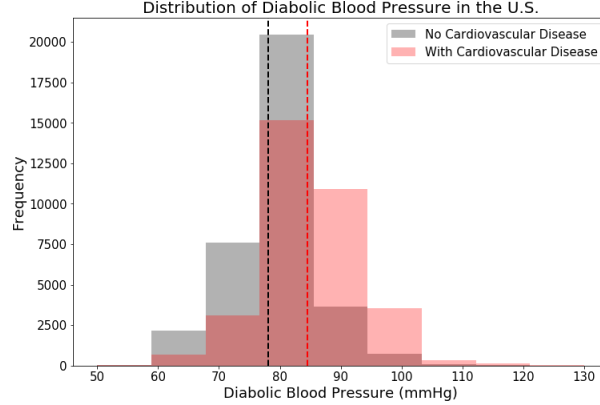
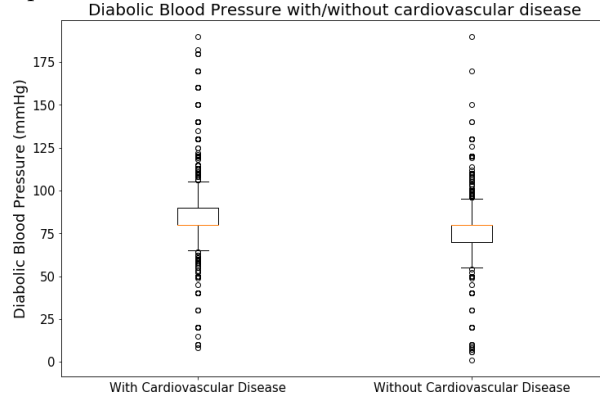


Figure 5: Grouped Boxplot of Diabolic Blood Pressure with and without cardiovascular disease



The average diabolic blood pressure of Americans with cardiovascular disease was 84.52 mmHg and the average diabolic blood pressure of Americans without cardiovascular disease was 78.11 mmHg which causes the average difference to be 6.40 mmHg.

This analysis demonstrates that average diabolic blood pressure of a patient with cardiovascular disease is higher than average diabolic blood pressure of a patient without cardiovascular disease. This can also be observed in the graph 4 and 5, where the mean and median of diabolic blood pressure is higher in patients with cardiovascular disease than patients without cardiovascular disease.

Conclusion

The statistical analysis provided evidence that factors like systolic blood pressure and diabolic blood pressure are directly related to cardiovascular disease. From the analysis, it can be stated that systolic and diabolic blood pressure have a linear relationship, which could further be related to cardiovascular diseases in United States of America.

Citations

- 1:** “WHO Reveals Leading Causes of Death and Disability Worldwide: 2000-2019.” World Health Organization, World Health Organization, www.who.int/news/item/09-12-2020-who-reveals-leading-causes-of-death-and-disability-worldwide-2000-2019.
- 2:** “Underlying Cause of Death, 1999-2019 Request.” Centers for Disease Control and Prevention, Centers for Disease Control and Prevention, wonder.cdc.gov/ucd-icd10.html
- 3:** Ulianova, S. (2019, January 20). Cardiovascular Disease dataset. Retrieved December 14, 2020, from <https://www.kaggle.com/sulianova/cardiovascular-disease-dataset>