**Investigating Indicators of Heart Disease**

Table of Contents:

1. Introduction
2. Research Question
3. Smoking and Heart Disease Association
4. Mental Health and Heart Disease Association
5. Alcohol Consumption & Heart Disease Association.
6. BMI and Heart Disease Association

1. **Introduction**

Cardiovascular disease (CVD), commonly known as heart disease, remains a leading cause of mortality worldwide. Understanding the factors contributing to CVD is crucial for prevention and intervention strategies. This report investigates potential indicators of heart disease. Specifically, we examine the associations between heart disease and modifiable lifestyle factors, such as smoking, alcohol consumption, and physical activity, as well as health conditions like stroke, diabetes, and mental health. The roles of other indicators and genetics are also briefly addressed.

1. **Research Questions**

This investigation seeks to answer the following key research questions:

* What are the strongest indicators that contribute to heart disease, and how does each indicator contribute to the severity of the disease?
* Are genetics a strong indicator for heart disease?

1. **Smoking and Heart Disease Association**

3.1 Hypothesis

Does smoking contribute to Heart disease?

3.2 Statistical Analysis: Chi-Square Test

Due to the categorical nature of "Smoking" and "HeartDisease" variables, a chi-square test of independence was performed to assess the association between these two factors. This test evaluates whether the observed frequencies of smoking and heart disease occurrence differ significantly from the expected frequencies under the assumption of independence.

3.3 Results and Interpretation

The chi-square test yielded a Chi-squared statistic of 3713.0331 and a P-value of 0.000000.

* The extremely low p-value (0.000000) indicates that the observed association between smoking and heart disease is highly statistically significant. This means it is highly unlikely to observe this strong association if smoking and heart disease were truly independent.
* The large chi-square statistic (3713.0331) further supports the rejection of the null hypothesis, indicating a significant difference between observed and expected frequencies.

A graph with blue and orange bars

AI-generated content may be incorrect.

3.4 Conclusion

The results strongly suggest a statistically significant association between smoking and heart disease. The rejection of the null hypothesis confirms that smoking and heart disease are not independent. Individuals who smoke are more likely to have heart disease.

Important Note: While the chi-square test demonstrates a significant association, it does not establish causation. Other factors may influence both smoking and heart disease. Further research is needed to determine the causal relationship.

The analysis provides strong evidence of a statistically significant association between smoking and heart disease. However, it is crucial to remember that association does not imply causation.

1. **Mental Health and Heart Disease Association**

4.1 Hypothesis

DoesMental Health contribute to heart disease?

Given that "Mental Health" is a numerical variable with a non-normal distribution, and "Heart Disease" is a binary categorical variable, a non-parametric Mann-Whitney U test was employed to assess the potential association between these variables. This test is suitable for comparing the medians of two independent groups when the data does not meet the assumptions of parametric tests, such as normality.

4.2 Statistical Analysis: Mann-Whitney U Test

The Mann-Whitney U test, also known as the Wilcoxon rank-sum test, is a distribution-free test that examines whether there are statistically significant differences between the medians of two independent groups. In this context, it evaluates whether the distribution of "Mental Health" scores differs significantly between individuals with and without heart disease.

4.3 Results and Interpretation

The Mann-Whitney U test yielded a p-value of 0.0514. This p-value is slightly above the conventional significance level of 0.05. See Figure 1 in Appendix for visualization.

* Borderline Significance:
  + A p-value of 0.0514 indicates borderline statistical significance. While it is close to the 0.05 threshold, it does not meet the standard criteria for rejecting the null hypothesis at this level.
  + This suggests that there is weak evidence against the null hypothesis, which states that there is no difference in the median "Mental Health" scores between individuals with and without heart disease.
* Weak Evidence of Association:
  + The results provide weak evidence of a potential association between mental health and heart disease.
  + It suggests that there might be a subtle difference in "MentalHealth" medians between the two groups, but the effect is not strong enough to be considered statistically significant at the 0.05 level.

A graph of a number of blue squares

AI-generated content may be incorrect.

4.4 Conclusion

The Mann-Whitney U test results suggest that there is weak evidence of a relationship between mental health and heart disease. While the p-value is close to the conventional significance level, it does not provide strong enough evidence to reject the null hypothesis at the 0.05 level. Further research, including larger sample sizes and the analysis of other relevant variables, is needed to explore this potential association in more detail.

1. **Alcohol Consumption & Heart Disease Association.**

5.1 Hypothesis:

Is there an association between alcohol consumption & heart disease?

The objective of this study was to examine the relationship between alcohol consumption and the incidence of heart disease. The null hypothesis (H₀) posits that there is no significant difference in the occurrence of heart disease between individuals who consume alcohol and those who do not. The alternative hypothesis (H₁) asserts that there is a significant difference in the occurrence of heart disease between these two groups.

5.2 Statistical Analysis

To test the hypothesis, two statistical methods were employed: the independent two-sample t-test and the Chi-Square test of independence.

* Independent Two-Sample T-Test: This test was used to compare the mean occurrence of heart disease between alcohol drinkers and non-drinkers. The t-test is appropriate for comparing the means of two independent groups when the data follows a normal distribution. The primary objective was to determine whether the average rates of heart disease differ significantly between the two groups.
* Chi-Square Test of Independence: A contingency table was constructed to assess the association between alcohol consumption and heart disease occurrence. The Chi-Square test was applied to evaluate whether alcohol consumption is independent of heart disease, by comparing the observed frequencies of heart disease within the two alcohol consumption categories against the expected frequencies under the null hypothesis of no association.

5.3 Results & Interpretation

* T-Test Results: The t-statistic was calculated to be -22.31, with a p-value of 2.64e-109. Since this p-value is substantially smaller than the conventional significance level of 0.05, the null hypothesis was rejected. This finding indicates that there is a statistically significant difference in the occurrence of heart disease between alcohol drinkers and non-drinkers. The negative t-statistic suggests that alcohol drinkers have a lower incidence of heart disease compared to non-drinkers.
* Chi-Square Test Results: The Chi-Square statistic was computed to be 328.65, with a p-value of 1.89e-73. As this p-value is also significantly smaller than 0.05, we reject the null hypothesis. The Chi-Square test thus confirms a significant association between alcohol consumption and heart disease. The discrepancy between the observed and expected frequencies further supports the idea that the variables—alcohol consumption and heart disease—are not independent of each other.

A graph of alcohol drinking

AI-generated content may be incorrect.

5.4 Conclusion

Both the independent t-test and Chi-Square test yield compelling evidence to reject the null hypothesis, indicating a significant relationship between alcohol consumption and the occurrence of heart disease. The t-test suggests that alcohol drinkers exhibit a lower incidence of heart disease compared to non-drinkers, while the Chi-Square test confirms that alcohol consumption is associated with heart disease occurrence. While these statistical findings establish a significant association, it is important to note that correlation does not imply causation. Additional research, incorporating potential confounding factors and more advanced statistical modeling, would be required to ascertain whether alcohol consumption directly influences the incidence of heart disease.

1. **BMI and Heart Disease Association**

6.1 Hypothesis

Higher BMI is Correlated with Heart Disease

6.2 Statistical Analysis: Spearman’s Coefficient & Mann-Whitney U Test

The BMI data was shown to be skewed using both histogram and QQ-plots. Thus, two non-parametric tests were explored.

6.3 Results and Interpretation

Both the Spearman’s Coefficient & Mann-Whitney U Test revealed extremely low p-values (3.7 e-231 and 9.0 e-231). Plotting the BMI data group by the presence/absence of Heart Disease (Yes/No) supports by visually showing the range of BMI is similar for both those with Heart Disease (Yes) and without Heart Disease (No).

A graph with black and blue lines

AI-generated content may be incorrect.

6.4 Conclusion

This dataset suggests no correlation between a person’s BMI and whether or not they have heart disease.