Final Report: Kaiden Purin

Background/Motivation for the Study

Smoking is a commonly thought of risk factor associated with hypertension (Erhardt), however, much of the research on the subject has been inconsistent in showing such (Sohn). While there are studies that have shown acute increases in blood pressure when smoking(Primatesta), those results haven't seemed to transfer to medically defined hypertension. Because of these inconsistent results I will compare hypertension between current smokers and non-smokers within the NHANES dataset to see how it compares against the results of other studies.

Research Question and Hypothesis

Within the NHANES data set is there a difference in the prevalence of hypertension between those who smoke and those who do not? Average systolic and diastolic blood pressures will also be compared to see if either group correlates with a higher level of each. We would expect to see higher prevalence of hypertension and higher average systolic and diastolic levels in the smoking group. However, due to the results of other studies I would be inclined to believe that the results would not be able to show that there is a direct correlation between smoking and hypertension, systolic blood pressure, or diastolic blood pressure. I anticipate that my results will show that regardless of the established idea that there is an association, the data does not necessarily align with it. The null and alternative hypotheses are as follows.

- The prevalence of hypertension is equal between smokers and non-smokers.
  - $H_0: P_{\text{smokerHypertension}} = P_{\text{non-smokerHypertension}}$
  - $H_A: P_{\text{smokerHypertension}} \neq P_{\text{non-smokerHypertension}}$
- There is no difference between measured systolic blood pressure between smokers and non-smokers.
  - $H_0: \mu_{\text{smokerSystolic}} = \mu_{\text{non-smokerSystolic}}$
  - $H_A: \mu_{\text{smokerSystolic}} \neq \mu_{\text{non-smokerSystolic}}$
- There is no difference between measured diastolic blood pressure between smokers and non-smokers.
  - $H_0: \mu_{\text{smokerDiastolic}} = \mu_{\text{non-smokerDiastolic}}$
  - $H_A: \mu_{\text{smokerDiastolic}} \neq \mu_{\text{non-smokerDiastolic}}$

Data Description and Exploratory Data Analysis

The initial data used is the NHANES dataset containing 76 variables with 10,000 observations. The variables being targeted are the average systolic (BPSysAve) and diastolic(BPDiaAve) measurements, as well as a new variable(hypertension) for if they do or do not have hypertension determined by the average systolic/diastolic compared against the current medically defined hypertension threshold of >= 130 mmHg for systolic and >= 80 mmHg for diastolic (Whelton et al.). The final variable is the SmokeNow column giving Yes or No for currently smoking tobacco. Omission of NA reported data in the SmokeNow and hypertension variables resulted in the removal of 6,892 observations.

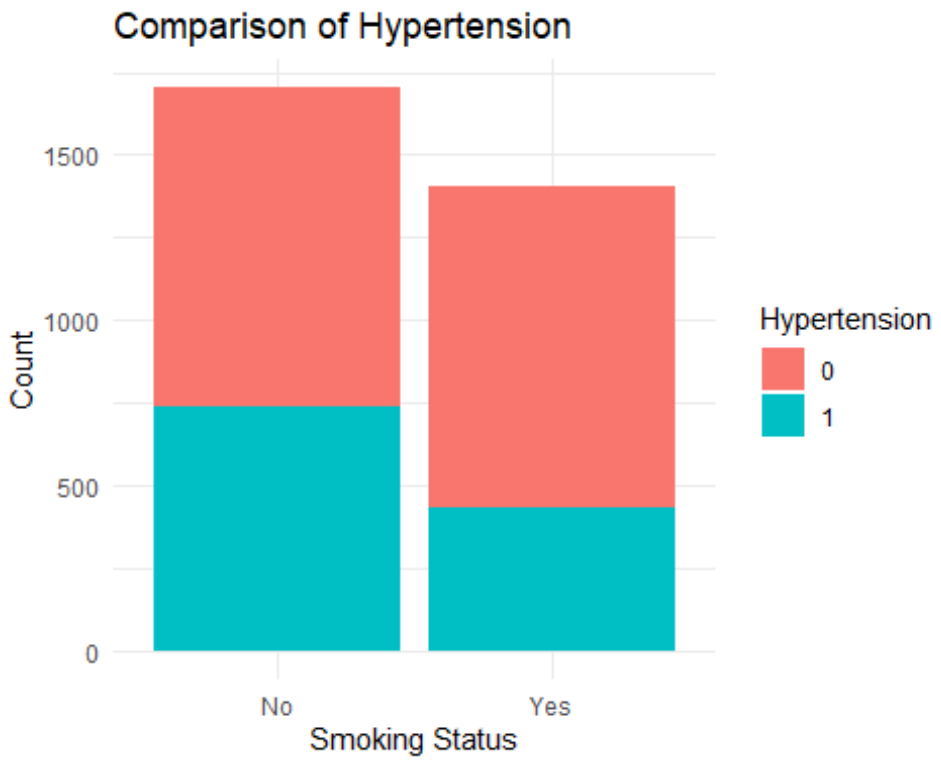
SmokeNow	hypertension	count
No	0	964
No	1	740
Yes	0	973
Yes	1	431

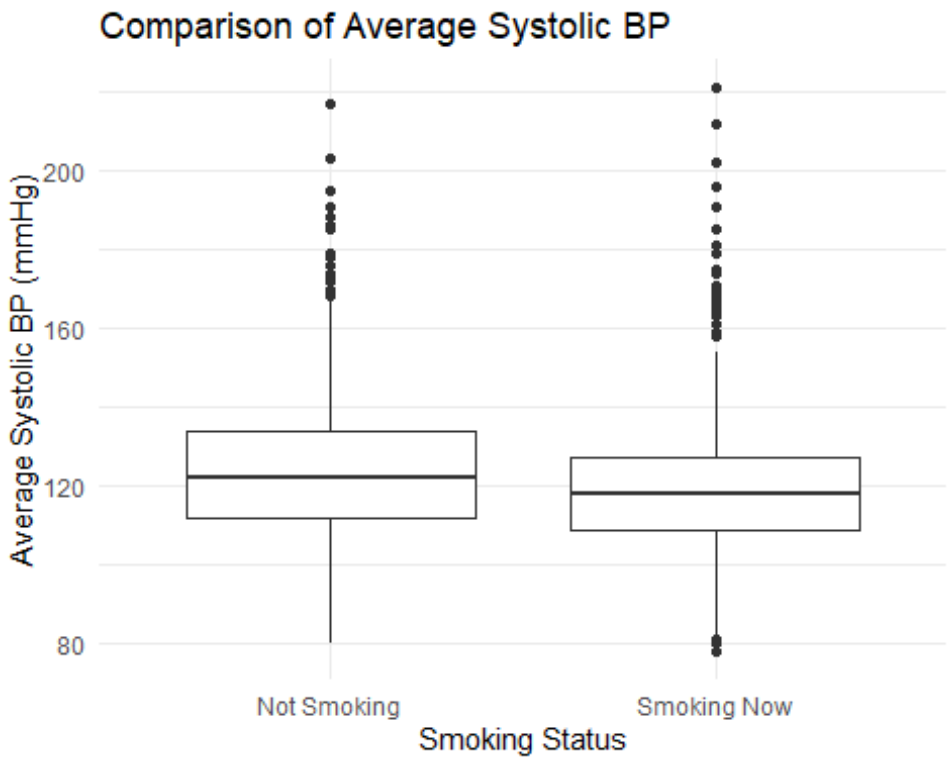
Group	Mean	SD
Smoking Group	0.3069801	0.4614054
Non-smoking Group	0.4342723	0.4958065

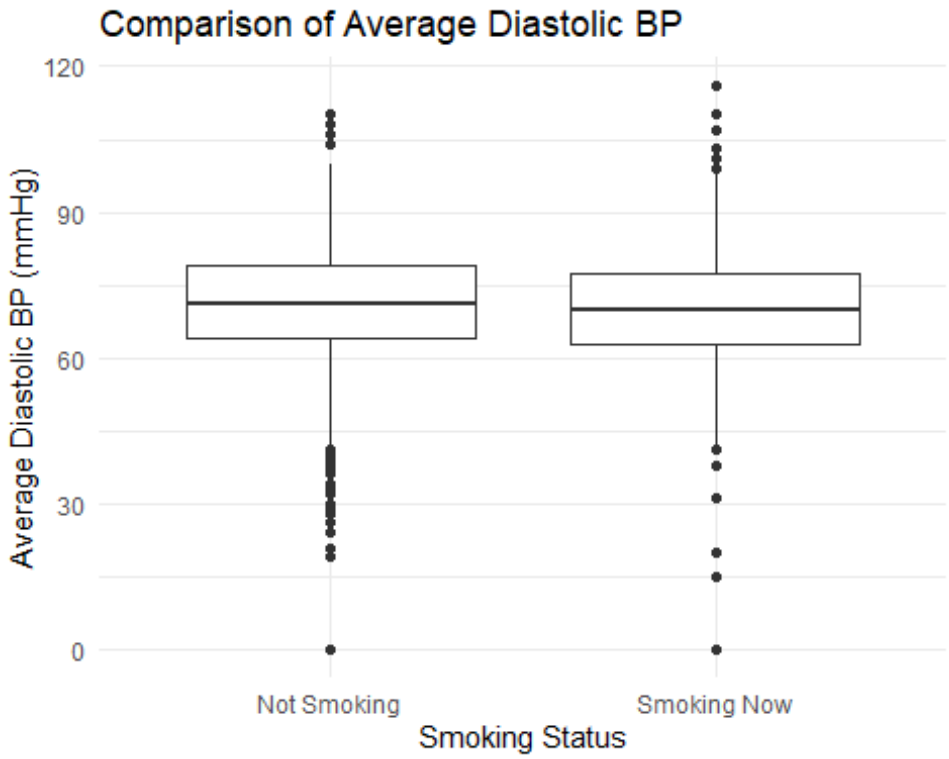
```
  SmokeNow min  Q1 median  Q3 max     mean       sd    n missing
1       No  80 112    122 134 217 124.3838 18.19431 1704       0
2      Yes  78 109    118 127 221 120.1197 17.37691 1404       0

  SmokeNow min Q1 median    Q3 max     mean       sd    n missing
1       No   0 64     71 79.00 110 70.55516 12.82171 1704       0
2      Yes   0 63     70 77.25 116 70.06553 12.28336 1404       0
```

There are a total of 3,108 observations in the sample with 1,704 non-smokers of which 740 have hypertension and 1404 smokers with 431 having hypertension. Within the observations that smoke the mean of those with hypertension is 0.31 and the standard deviation is 0.46. The non-smoking observations have a mean of 0.43 and a standard deviation of 0.50. When comparing mean systolic blood pressure the non-smoking group has a mean of 124.38 mmHg and the smoking group has a mean of 120.12 mmHg. For mean diastolic levels there is minimal difference with the non-smoking group at 70.56 mmHg and the smoking group measured at 70.06 mmHg.


Comparison of Hypertension
1500
1000
500
0
Count
No
Yes
Smoking Status
Hypertension
0
1


Comparison of Average Systolic BP
200
160
120
80
Average Systolic BP (mmHg)
Not Smoking
Smoking Now
Smoking Status



A stacked bar graph was utilized to visualize the difference in portion of hypertension between each group. In this graph we are able to see that there appears to be a lower prevalence of hypertension within the smoking group. This would support the idea that the results are inconsistent when relating smoking to hypertension as it would appear to be contradictory to the idea that smoking is a major risk-factor for hypertension. In the box plots we are able to visualize that the IQR of mean systolic blood pressure is lower in those that do smoke when compared to those that do not. The visualization for mean diastolic blood pressure appears to indicate that the IQR for both groups is relatively equivalent with the non-smoking group having far more outliers in the lower end.

Analysis

```
    Pearson's Chi-squared test with Yates' continuity correction

data:  contingency_table
X-squared = 52.576, df = 1, p-value = 4.138e-13
```

```
    Welch Two Sample t-test

data:  non_smoke_group$BPSysAve and smoke_group$BPSysAve
t = 6.6649, df = 3039.3, p-value = 3.133e-11
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 3.009670 5.518619
sample estimates:
mean of x mean of y
 124.3838  120.1197


    Welch Two Sample t-test

data:  non_smoke_group$BPDiaAve and smoke_group$BPDiaAve
t = 1.0842, df = 3036.6, p-value = 0.2783
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.3958336  1.3751081
sample estimates:
mean of x mean of y
 70.55516  70.06553
```

Because we are comparing the proportion of hypertension between the smokers and non-smokers a chi-squared test was utilized. Based on the p-value we are able to reject the null hypothesis in favor of the alternative hypothesis and would conclude that there is a statistically significant difference between group proportions. Furthermore, the x^2 value indicates that there is a strong association between smoking status and hypertension within this data set.

For comparing systolic/diastolic blood pressure t-tests were utilized to determine if there is a statistically significant difference in means between the groups. Through the systolic t-test we have enough evidence to reject the null hypothesis that there is no difference in mean systolic levels based upon the p-value as well as the confidence interval. We are 95% confident that there is a statistically significant difference between the groups. Looking at the means of each group and the confidence interval we can see that, on average, the non-smoking group has a higher systolic blood pressure compared to the smoking group. The diastolic t-test resulted in p-value higher than the generally accepted $\alpha(0.05)$ as well as a confidence interval that contains 0. Because of this we fail to reject the null hypothesis and conclude with 95% confidence that the mean diastolic blood pressure between the groups is the same on average.

Conclusions

In this analysis we were attempting to determine if there is a discernible difference in the prevalence of hypertension between those who do and do not smoke. Chi-squared tests, t-tests, and visualizations were used for the comparison. Through the results of these tests we are able to conclude that there is a difference when comparing hypertension prevalence and systolic blood pressure within the groups. Contrary to the general idea of association between hypertension and smoking, the non-smoking group displayed a higher prevalence of hypertension and higher mean systolic blood pressure on average. These results are visually apparent in the stacked bar graph(hypertension) and the box plot(systolic) as well as statistically shown in the chi-squared test and t-test. The comparison of diastolic blood pressure however showed no statistically significant difference according to the two sample t-tests which could be inferred from the box plot.

The final results of this analysis align well with the outcome I expected based upon previous studies conducted on the topic. The results are inconsistent with the idea that there is a positive association between smoking and elevated blood pressure and if anything, infers the opposite. This shows that there are potentially several other important factors to consider when making this comparison in future research.

Works Cited

Erhardt, Leif. "Cigarette Smoking: An Undertreated Risk Factor for Cardiovascular Disease." *Atherosclerosis*, vol. 205, no. 1, July 2009, pp. 23–32, https://doi.org/10.1016/j.atherosclerosis.2009.01.007.

Primatesta, Paola, et al. "Association between Smoking and Blood Pressure." *Hypertension*, vol. 37, no. 2, Feb. 2001, pp. 187–193, www.ahajournals.org/doi/full/10.1161/01.HYP.37.2.187, https://doi.org/10.1161/01.hyp.37.2.187.

Sohn, Kitae. "Relationship of Smoking to Hypertension in a Developing Country." *Global Heart*, vol. 13, no. 4, Dec. 2018, pp. 285–292, https://doi.org/10.1016/j.gheart.2018.01.004.

Whelton, Paul K., et al. "Harmonization of the American College of Cardiology/American Heart Association and European Society of Cardiology/European Society of Hypertension Blood Pressure/Hypertension Guidelines: Comparisons, Reflections, and Recommendations." *Circulation*, vol. 146, no. 11, 13 Sept. 2022, pp. 868–877, https://doi.org/10.1161/circulationaha.121.054602.