**Introduction**

An analysis was conducted to analyze how prevalence of stroke in men and women is related to body mass index (BMI). The dataset used is from the Framingham Heart Study conducted between 1956 and 1968. The study initially enrolled 5,209 subjects, but the dataset contains data on 4,434 subjects. The purpose of the study was to track risk factors for cardiovascular disease such as blood pressure, lung function, smoking history, BMI, and stroke prevalence. The subjects were observed over time and regular checkups every 6 years allowed researchers to determine what factors contribute to certain cardiovascular issues.

Two categorical variables were chosen: Prevalence of Stroke (P: prevalent disease, NP: free of disease) which represents the likelihood of the subject suffering a stroke, and sex (male and female). One numerical variable was chosen, BMI (measured in weight in kilograms/height in meters squared), representing the subject’s body mass index at the time of each checkup. Higher BMI values typically correlate to less-healthy individuals.

The hypothesis for the main effect of prevalence of stroke on SMI:

***H0: µP-BMI = µNP-BMI***

***Ha: µP-BMI ≠ µNP-BMI***

The hypothesis for the main effects of sex on BMI:

***H0: µmale-BMI = µfemale-BMI***

***Ha: µmale-BMI ≠ µfemale-BMI***

The purpose of this analysis is to determine if BMIs in males and females with a prevalence of stroke is higher than those who have no prevalence.

**Methods**

The analysis performed was a two-ANOVA unbalanced factorial design to test the main effects of sex and stroke prevalence on BMI. The reasoning for choosing this test is that there are two categorical independent variables, sex and stroke prevalence, and one numerical dependent variable, BMI. There were no repeated measures and the unbalanced design was chosen as there are not equal numbers of males and females. The data is from the first checkup period, so period = 1, with the analysis being performed in R at a 5% significance level. Also, the indicator variables for sex and stroke prevalence were instead given character values. 1 was changed to male, 2 to female, and for stroke prevalence 0 was changed to NP (free of disease) and 1 changed to P (prevalence of disease).

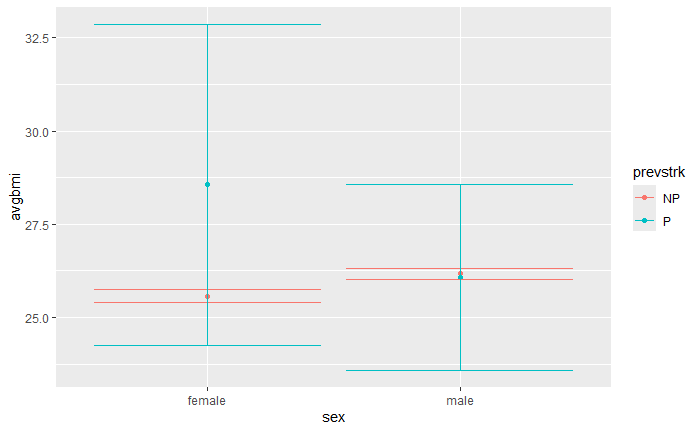
**Results**

The ANOVA test produced the following results.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | DF | Sum Sq | F-value | P-value |
| Sex | 1 | 362 | 21.65 | 3.37e-6 |
| Prevalence of Stroke | 1 | 81 | 4.86 | 0.03 |
| Sex:PrevStroke | 1 | 69 | 4.12 | 0.04 |
| Residuals | 4411 | 73753 | - | - |

Testing for the main effect of sex on BMI, the F-statistic was 21.65 with p-value of effectively zero. This means that we can reject the null hypothesis and conclude that there is a significant mean difference in BMI between males and females. Next, testing for the effect of stroke prevalence on BMI, the F-statistic is 4.86 and p-value of 0.03. Therefore, we can also reject the null hypothesis and conclude there is a significant mean difference in BMI between those free of disease and those with a prevalence of stroke-causing diseases. Looking at the interaction between sex and prevalence of stroke, the F-statistic was 4.12 and p-value 0.04, indicating that there is a significant interaction between the sex of the subject and their risk of stroke.

The graph below shows average BMI with error bars.



The graph interestingly shows that the average BMI for males is very similar, regardless of stroke prevalence, while the average BMI for females is much higher for females with a prevalence of stroke.

**Conclusion**

There are many risk factors for stroke, such as smoking, high cholesterol, diabetes, and cardiovascular disease. Many of these risk factors are caused, or exacerbated, by a high BMI. People can have genetic factors that increase their risk of stroke despite otherwise being perfectly healthy individuals. I was interested in seeing how the risk of stroke was related to BMI, and if it could be shown that individuals with higher stroke risk typically also had higher BMIs. Gender is also a contributing factor to the prevalence of stroke and thus it was important to look at each gender individually.

Interestingly, the average BMI for males is very similar for individuals with a prevalence for stroke and those with no prevalence. In fact, the average BMI for males without any prevalence is slightly higher than those who do have prevalent risk. This would contradict the conventional wisdom that individuals with a higher BMI are more at-risk for strokes than those with lower BMIs. However, the variation of BMI for males with a prevalence for stroke was much larger than that of non-prevalent males, indicating that the factors that contribute to stroke risk and BMI are individually different. For females, the average BMI for those at risk is much higher than those without risk, which leads to the notion that higher-BMI females typically have a high prevalence for stroke-causing diseases. It would seem that the majority of the interaction between sex and stroke prevalence comes from female side. Another interesting finding is the variation of BMI for prevalent risk individuals. For no prevalence of disease, the BMI variation is quite low, but there’s much more uncertainty when the individual is at-risk. Lower BMI is a good indicator that the person is in good health, so those seemingly healthier individuals having more prevalent stroke risk is somewhat surprising.

The limitations of the study are that there are many confounding factors for stroke prevalence as well as BMI. Healthy (i.e. low BMI) individuals could have genetic factors that increase their chance of stroke-causing diseases. Other factors, such as the subject being diabetic, could cause a higher BMI and therefore increase their risk of stroke-causing disease. A more in-depth analysis that includes more variables would be necessary in order to fully understand if individuals with a higher BMI are generally more at risk for strokes. Controlling for factors such as diet and exercise, and whether the prevalence for stroke-causing disease is genetic, would contribute to an improved analysis.

In conclusion, while this analysis succeeding in showing that individuals with a prevalence of stroke-causing diseases typically had a higher BMI than those with no prevalence, deeper analysis is necessary to understand the link between stroke prevalence and BMI.