Kathryn Howe

DSC530-T301

Prof. Cary Jim

10 August 2024

**Final Project: Investigation of Alzheimer's Disease**

My analysis for this final project explored potential correlation between various lifestyles and demographics of older patients and an Alzheimer's disease diagnosis. This research attempts to identify a correlating variable to assist those with potential diagnosis, or determine if certain demographics or lifestyle choices predispose certain people to Alzheimer's disease. The data for this research was taken from Kaggle.com and the uploader provided significant amounts of background information and information regarding the variables that indicate that this was a credible source. This is the link to the dataset for potential future analysis: Alzheimer's disease dataset.

The variables I pulled for my analysis were the following. Where the range is zero to one, zero indicates a no and one indicates a yes.

| Variable | Range |
|---|---|
| Diagnosis | 0-1 |
| FamilyHistoryAlzheimers | 0-1 |
| Age | 60-90 |
| SleepQuality | 4-10 |
| HeadInjury | 0-1 |
| DietQuality | 0-10 |
| Mini-Mental State Examination (MMSE) | 0-30 |

Overall, my analysis provided little to no correlation between the variables and a positive or negative diagnosis. Nearly every diagram appeared the same between those diagnosed and those undiagnosed. There was a lot of information provided in the dataset. Something that I may have missed or that could have had a correlation with the diagnosis could have just been a variable that I didn't test. There were over 30 different variables and questions presented to the patients and I only examined 7, so a potential correlation could be within one of those other variables.

I feel as though, any variable that could be needed is present in this dataset so any future questions just need more analysis with the other variables present. Some assumptions I made or that could have been made with the data was that there was going to be a correlation between variables. So I was looking for any correlation, even a slight one. The OLS regression results indicated a slight relationship between SleepQuality and Diagnosis, however considering the other analysis and graphs produced, this relationship is likely coincidental, as all the other graphs and analysis indicate a uniform distribution where the variables do not have an impact on one another.

A challenge I faced was cleaning the data, as it required me to learn new skills within Excel. I also have some difficulty reading the OLS results, but I was able to determine that through the readings. Overall, I don't fully understand how there can be little to no correlation, however it might just be true that Alzheimer's is a disease that randomly afflicts people.