COMP20008 Assignment 2

Introduction

Nowadays, words such as "mental health" and "psychological consultation" have been understood and recognized by more and more people. This is related to the acceleration of modern society, the explosion of information, and the increase in pressure. In addition to paying attention to physical health, people pay attention to their own mental health, and to improve the quality of life as a whole has become a very important part of modern life. Mental health takes its toll on the society, it costs human lives and reduces people's ability to work even live, which leads the panic of the whole society. Studying this topic would help us to find out the extent that the mental problem affects our daily life. Therefore, this project aims to find the factors that may affect mental health, for example, each person's lifestyle, living environment, economic income and so on.

This analysis offers linear approaches for our research question of whether variables such as income, rent paid for each person and vehicle hold for each household have positive or negative relation with respect to people's mental health. By analyzing the data of each person's income, household's vehicle hold, rent paid from different areas in Australia, the factors that associated with their mental health conditions can be shown by concluding the statistical results.

The remainder of the report was organized as follows. In the method section, we first explain the process of data wrangling and the methods of our analyses by interpret the effect of each variable. Then, we give the statistical results of our analyses in the result section. Finally, we summarize the results of our study and discuss its limitations.

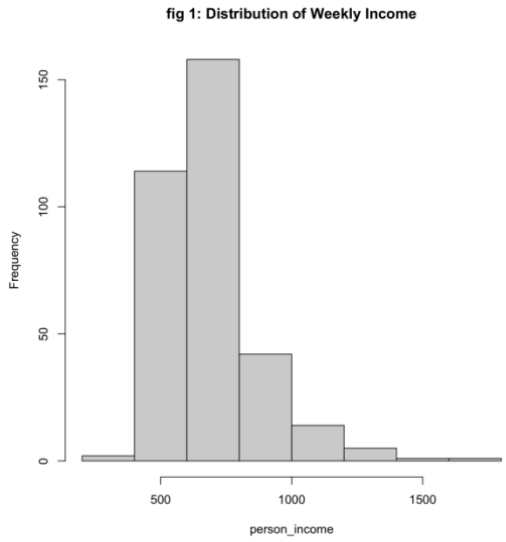
Data Summary

We collect data from two source, the first one is Australia 2016 census data by Australia statistical Area Level 3 and the second one is Mental health services in Australia: Medicare-subsided mental health-specific services by Australia statistical Area Level 3. For the census

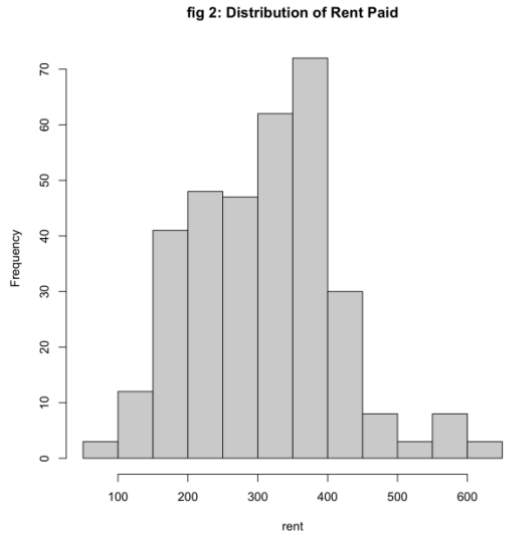
data we retrieve three list of data: median personal weekly income, median rent in this area and how many vehicles hold by each family on average in this area. For the second data, we retrieve the number of patients seeking help in the mental health related services in that area. In order to make the data comparable among different area, we divide the number of patients seeking help by the total population in that area to obtain the ratio of people seeking help in mental related services. We found that there are some regions did not record the number of patients seeking help, therefore we decide to remove the corresponding row. Finally, we obtain 337 data from different region.

These two datasets are from Australian bureau of statistics and Australian institute of health and welfare respectively. Both of the website was running by government officially. The second datasets collect the data by the record of medicare, and the census data is collected with a high respond rate. Therefore, we believe the data are reliable.

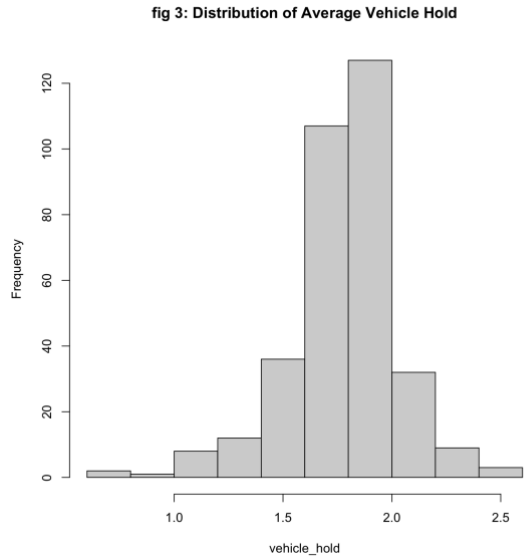
We first visualize the distribution of each variable to see their overall trend. For the distribution of weekly median income of each person shown in figure 1, we could see that most people receive 500 AUD to 600 AUD income weekly, for more than 70% area of these selected regions.



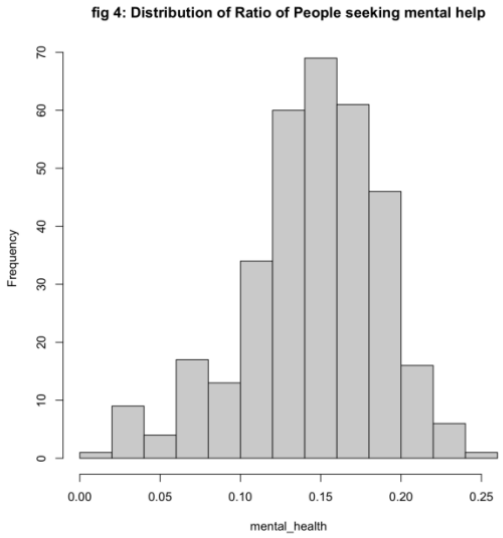
For the distribution of the median price for house rent per week, we could see that most rent paid interval of Australian is between 200 AUD to 400 AUD per week.



For the distribution of how many vehicles each household owns on average, we could see that most household at least have one vehicle.



For the distribution of the ratio of the people going to see psychologists, we could see that the data are normally distributed with the ratio of people seeking for psychotherapy most lying between 1.5 and 2.0.



As for plotting the histograms, the results indicated that only the plot for the weekly median income data is right-skewed. For the rest three variables i.e., price for house rent; average cars owned per person; and the ratio of people seeking for psychological help, histograms for these variables seem to have normally distributed pattern.

Methods

We use R rather than Python to analyze our research question since R is open-source software and simple in build statistical models which help us quickly understand the relationship among the factors while Python is more efficient on data wrangling.

At first, for each collected variables, we used the histogram function in order to see the distributions of those variables. Secondly, a simple linear model was chosen to fit the data. We used the ratio of people going to psychologist to be the response variable. As for the explanatory variables, we chose price for house rent, average cars owned per person and weekly median income of each person. And then by checking the summary and the ANOVA test result of this saturated model (rent + car + income), we removed the non-significant terms which would lead us to the reduced model (rent + income). We would then still check the summary and ANOVA test result for the new model. The final step is to do the ANOVA test between the saturated

model and the reduced model.

Results

As for model fitting, both the summary and the ANOVA result are consistent for the saturated model showing that the term "car" (i.e., average cars owned per person) is insignificant. So, the reduced model is the model without the "car" term, according to the ANOVA test, both "rent" and "income" has significantly small p-value. This means that both price for weekly house rent and weekly median income of each person have effect on the ratio of people going to see psychologist. Last but not least, by comparing the saturated and reduced model, the ANOVA test result shows that the reduced model is better.

As for the visualizations of the response variable mental health versus other factors, which shown in the following figures. We could see that the ratio of mental health of each area in Australia has negative relationship with the weekly income of each person, which means as the weekly income of a person increases, this probability of this person to see a psychologist decrease. Also, the rent paid shows positive relationship with the ratio of mental health while the average vehicle hold performs negative correlation with the ratio of mental health. In other words, as the weekly rent paid increase, people might have higher probabilities to see the psychologists. By contrast, the more a household hold vehicle, the lower probability a person to seek the psychotherapy.

fig 4: Mental Health vs Weekly Income

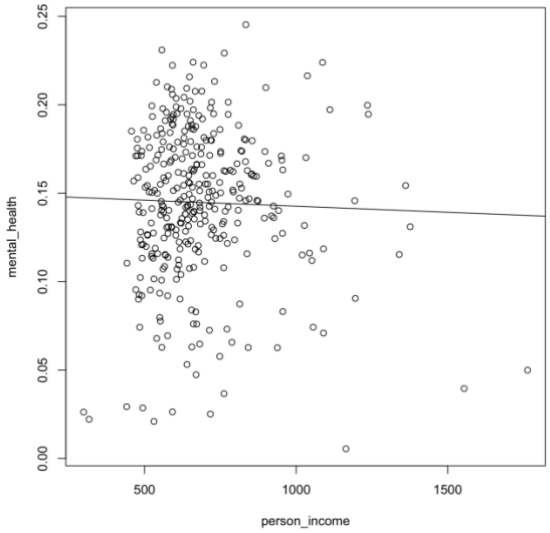


fig 5: Mental Health vs Rent Paid

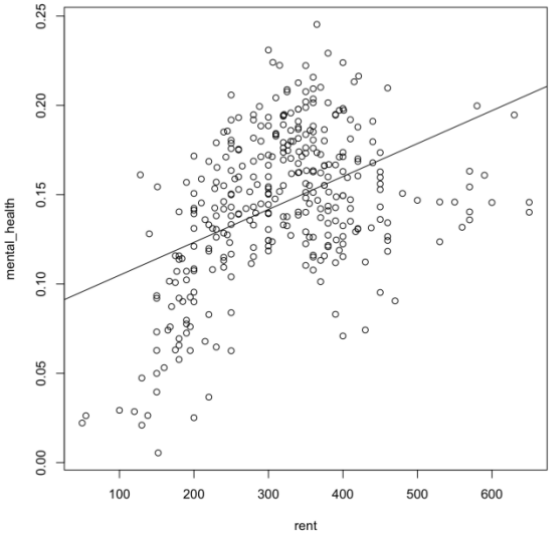
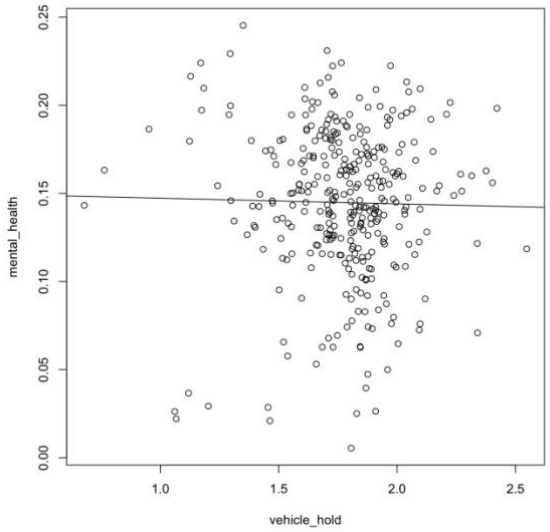


fig 6: Mental Health vs Vehicle Hold



Limitation

The ratio of people seeking help for mental related service may not represent the actual ratio of people having mental issue. The pool people may not want to see a psychologist or related help even they have mental issue because the expense of this kind of service is high. Moreover, people may not want to see a psychologist just because they did not realize this is a good way to handle the problem or they may think seeing a psychologist is shameful. To resolve this limitation, it will be a good idea to use other mental health report instead of record of medicare. There are other factor having stronger correlation with mental issue ratio which may be more useful and interesting to analysis, we choose these data (personal income, household vehicle holds on average and weekly rent in this area) because this is the data we can obtain and believe to be reliable. Last but not the least, it will be better if we use the data directly link a single person basic information and the mental health instead of analyzing the data for a whole region.

Conclusion

The data analysis method helps to explain why some of the region has higher rate of mental health problem over Australia, we found that there is indeed some correlation between the mental health issue and personal income and weekly rent they pay.

Datasets

2016 and historical Census data.

https://datapacks.censusdata.abs.gov.au/datapacks/

Mental health services in Australia: Medicare-subsidised mental health-specific services by Australia statistical Area Level 3

https://www.aihw.gov.au/reports-data/health-welfare-services/mental-health-services/data