DSCT Assignment 1

Zheng Yi Tao

Questions

1. Yes, there are missing values in the dataset. According to the codebook, missing values are denoted as the following:
   1. "8", "98" or "999998" as "Don't Know"
   2. "9", "99" or "999999" as "No answer"
   3. "0" as "Inapplicable"
2. Solution shown in the code.
3. Solution shown in the code.
4. Solution shown in the code.
5. Using summary and describe function from R, an elaborated summaries has been generated. From the observed summary, the mean of per-capital income is much lower than the mean of family income. At the same time, the standard deviation of the per-capital income is also lower, showing that there is less variance in the data as compared to family income.
6. The answer is no. According to the bar plot, there are more males who stated they are “Very Happy” than females and that more females stated that they are “Not too happy” than males. However, there are actually more females who say that they are “Pretty happy” than males. This does not however negate the answer because females are only approximately 4% more under “Pretty happy” category while male have approximately 7% more than female under “Very happy” category.

But if we were to analyse the data deeper, using Chisq.test() function in R programming, we have seen that with the chi-square value at 1.4826, degree of freedom at 2, we get the p-value of 0.4765 which is greater than the significance value of 0.05 which means that we cannot reject the null hypothesis that the gender variables are independent of the happiness variables. Which says that there is a chance that this test is not valid due to the fact that gender and happiness variables in this data has no effect on one another.

1. Yes, younger people are happier. A bar plot was generated with the ages divided into 3 categories – “Young”, “Middle” and “Old”. Where “Young” is between 0-25 years old, “Middle” is between 26-50 years old and “Old” is 51 and above. According to the bar plot, young people are approximately 4% and 7% higher than middle and old people respectively in the “Pretty happy” category. Old people on the other hand are 5% and 3% higher than young and middle aged people respectively in the “Not too happy” category. However, there is a small case of old people being approximately 2% and 0.01% higher than young and middle aged people respectively in the “Very happy” category. But majority as seen from the percentages shows that the younger the person is, the happier they are.
2. The answer is yes. As consistently shown in the aggregate function when calculating the median of the people’s income in relation to their happiness. Being “Not too happy” has the lowest median income of $13475, “Pretty happy” has the second highest income of $22050 and “Very happy” has the highest income of $26950.
3. The answer is yes. As seen from the aggregate function when calculating the median of the people’s income in relation to their health. From the lowest health of “Poor” category, the income is the lowest at $10412.5 and it progresses consistently through “Fair”, “Good” and “Excellent” with the income rises to $33075 at “Excellent” category.
4. The answer is yes. Even though the data for both happiness and marital status are categorical – which makes co-relation difficult to check – but using the summary on the linear model still managed to show the output. The p-value of the result showed that the relationship is highly significance and that they are indeed very dependent on each other as the p-value is much lesser than 0.05.
5. Yes, there are household income differences between races. Among the races, “Black” has the lowest income, while “Other” has the second highest income and “White” has the most income with almost twice as much as the “Black” race.