**Part I**

We expect there to be a relationship between these two variables since it would be easier for the poor to compare themselves with the rich population. For example, if everyone in the country is poor people might accept their situation as normal and do not realize that there is an alternative lifestyle. But if there is a big gap then the poor would see the lifestyle of the rich and this might cause them to feel worse about their own situation.

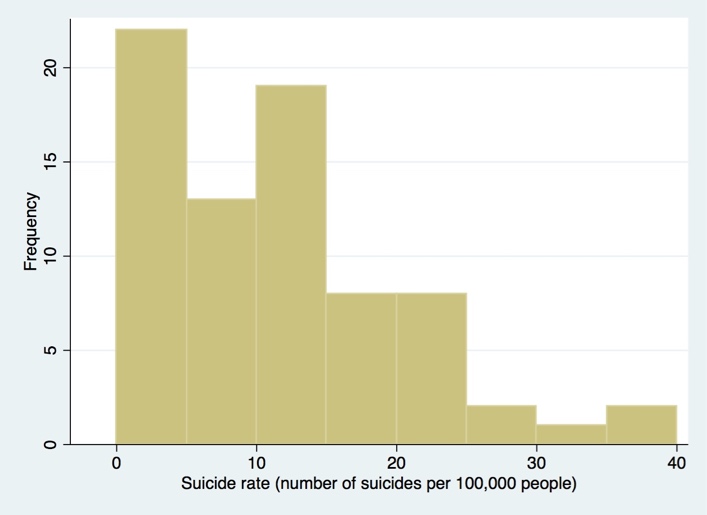
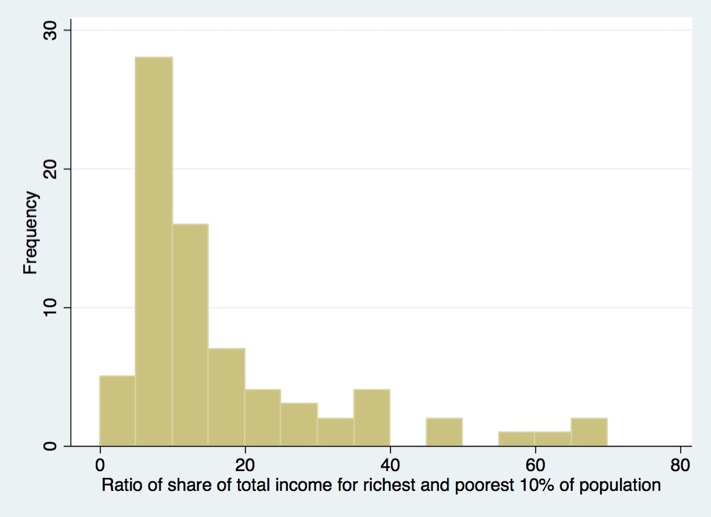
We expect the variables to be positively related, meaning that if income disparity is higher the suicide rate would also be higher. We expect the relationship to be modestly related. Because it could be the case that an overall rich country with higher income disparity have lower suicide rate than an overall poor country with lower income disparity. This means that it might be the case that the cause for suicide be poverty itself and not the income disparity. In this case we would not see a relationship between income disparity and suicide rate unless there is also a relationship between poverty and income disparity. It would make more sense that income disparity be the cause for higher suicide rates instead of the other way around. (there is no obvious explanation of why suicide rate would cause income disparity) Therefore the income disparity would be the independent variable and suicide rate would be the dependent variable.

**Part II**

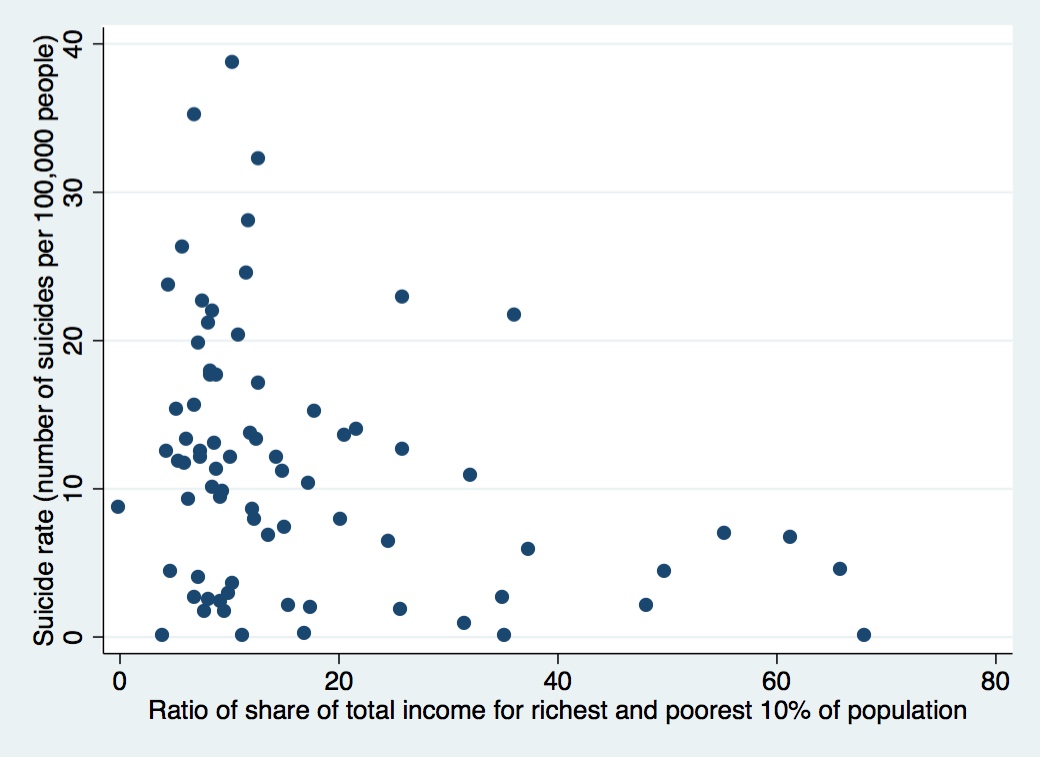
According to the variable descriptions, variable suicidert measures the number of suicides per 100,000 of population. This variable has an interval/ratio level of measurement. The variable richpoor is the ratio of income for 10% richest to the income of 10% poorest of the population. The larger the value for richpoor the higher income disparity in the country. This variable has an interval/ration level of measurement.

The distribution of the two variables in the dataset can be seen from the histograms below:

*histogram richpoor, frequency width (5)  
histogram suicidert, frequency width (5)*

  
Figure 1. Histograms for income disparity (left) and suicide rate (right)

As can be seen from the above graphs. The income disparity seems to follow a normal distribution skewed to the right with mean around 10. At the extreme this ratio goes as high as 70. The suicide rate seems to follow a similar looking normally distributed distribution skewed to the right with the mean around 10, the maximum at around 40 and the mode at 3. We can see the relative distribution of the two variables by looking at the scatterplot of the two:

  
Figure 2. Scatterplot of suicide rate vs income disparity

We can see from the above scatterplot that there seems to be a week negative relationship between the two variables.

**Part III**

No recoding of the variables is needed since our variables are not categorical. We could change the units in order to make the relationship in the graph more obvious but since Strata automatically scales the graph, this would not be needed. The only modification I made to the dataset was to remove the rows that did not have a valid value (empty cells) from our data set:

*drop if (mi(richpoor)) | (mi(suicidert))*

Since the scatterplot does not show a clear linear relationship (the relationship might be curvilinear) and since the data distributions, seen in the histograms of Figure 1, are not symmetrical, it would be better to use Spearman’s rho as the statistical measure of correlation between the two variables:

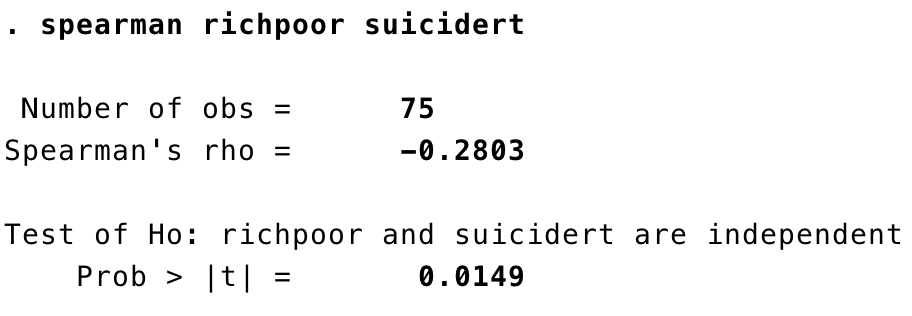


Figure 4 – Spearman’s rho and test of significance

We see that the Spearman’s rho measure shows a modest correlation in the negative direction. The test of significance shows a significance level of 1.5% which is statistically significant assuming alpha level of 5%.

**Part IV**

In part III we found that there seems to be a negative and modest relationship between suicide rate and income disparity. This means that as income disparity goes higher the suicide goes down. Therefore, our expectations was wrong and we realize that the correlation is in the opposite direction than what we first believed it to be. However, we cannot assume causation since as mentioned in Part I it could be the case that an overall rich country with higher income disparity have lower suicide rate than an overall poor country with lower income disparity. In this scenario country’s wealth could be the antecedent variable for the causal relationship between *richpoor* and *suicidert*, or it could be the cause of both, therefore creating a spurious relationship between the two. To establish whether or not there is a causal relationship between suicide rate and income disparity, and to figure out the effect and nature of this causal relationship, we would need to study the the correlation between all the variables that could affect suicide rate and suicide rate.