Introduction

The aim of my research is to find out if and which socioeconomic factors have the most significant relationship with COVID-19 contraction. I would like to see if some people in society were more likely to contact the dreaded virus because of factors which they could not control such as race, gender, and ethnicity. I also want to see how much someone's financial status might affect the possibility of contacting and even surviving the disease. If we know how this disease might affect different demographics, then the people in charge can act accordingly by making sure those at a higher risk are provided more resources.

I am using a dataset gathered mainly from the US CDC by Garcia (2020). It represents the different number of covid cases in the US according to race/ethnicity, gender, poverty, severe health conditions and Covid 19 mortality. Every entry is for each county in the country. Peters (2020) was interested in using similar data to create a susceptibility and vulnerability scale at the county level of the country: should different measures be adopted in different regions? Using a multivariate general linear model, he found that rural populations required different measures than large metropolitan areas. Patel et al (2020) also acknowledged that more factors must be considered when calculating the level of vulnerability of people to the virus. So, I have conducted my research with the following questions in mind:

1. Do areas with higher risk have larger populations?
2. Do counties with more impoverished people have more cases than counties with less?
3. Does more poverty mean more deaths?
4. Is there a relationship between number of cases and race?

Data

The dataset I am using shows the number of covid cases for each county in the US. It breaks down the total cases for each county according to many categories, but I will only be using the ones for people who died and people who are in poverty. Each row in the dataset represents data entries for each county in the US. The dataset also contains 3107 rows of data.

To describe the data, we first look at the histogram of people who got the disease and were, at the time, experiencing poverty.
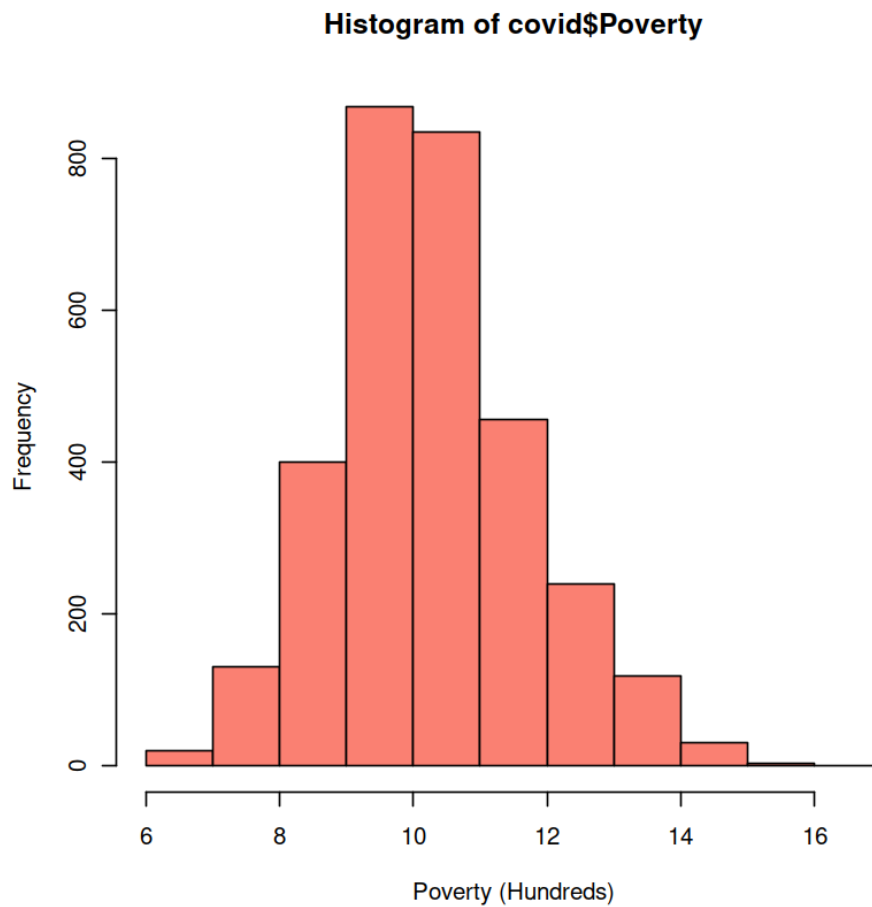
Histogram of covid$Poverty

Figure 1.

Figure 1 shows that the data is roughly symmetric and roughly shows the median number of people in poverty in the country to be 1000 people.

We also look at the histogram of people who got the disease and died from it.
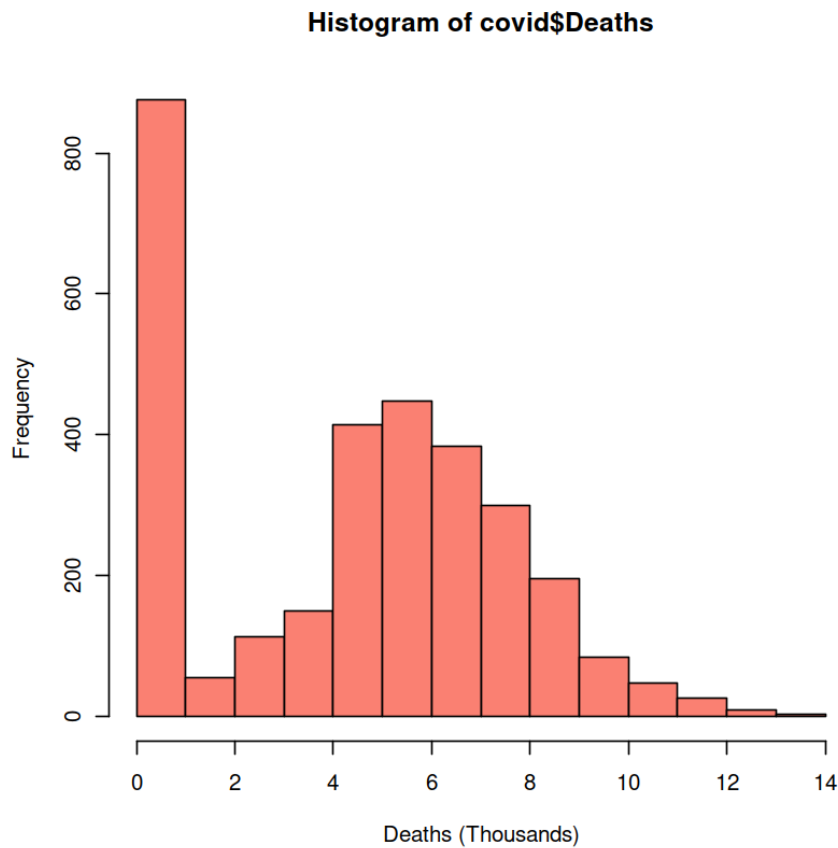
Histogram of covid$Deaths

Figure 2.

Figure 2 shows the data is skewed positively and is also bimodal as there is a distinctive difference between the beginning of the data and the rest.

Analysis

The first question I'm answering is "Do counties with more impoverished people have more cases than counties with less?". To do this, I decided to compare the number of cases for each county to the number of people in poverty in that county. I plotted these values and got the following result.
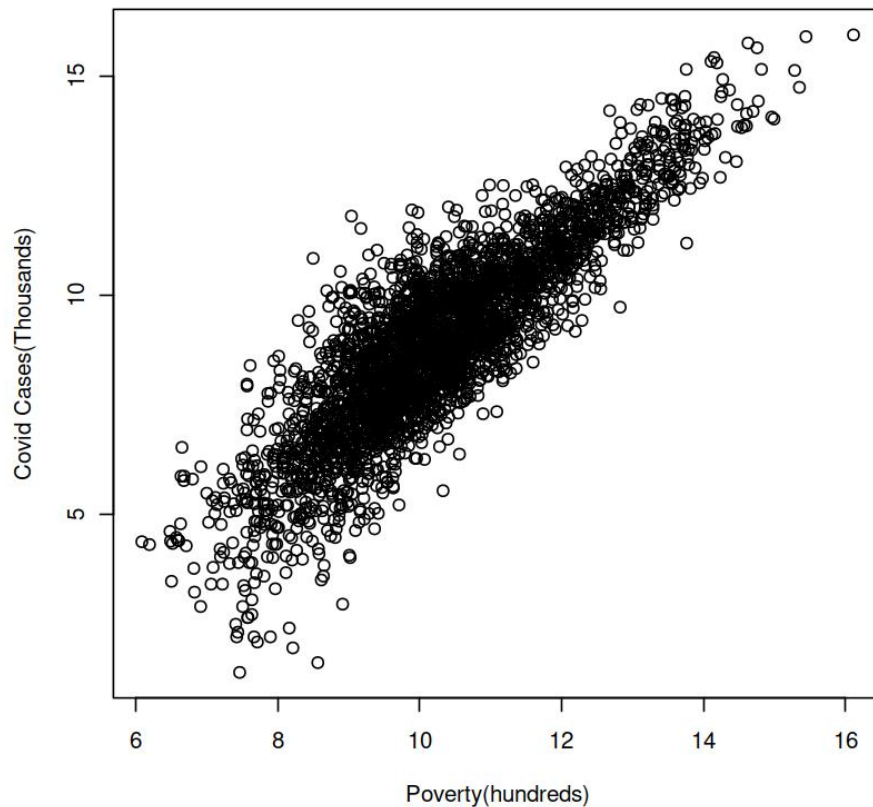
Figure 3.

Just by looking at the graph in figure 3, we can see that there is a positive correlation between the two values. As poverty increases the number of covid cases increases also. To confirm the significance of this graph, I did a Pearson's correlation test and got a p value of 2.2×10^{-16} which shows that it is statistically significant. I also found that the value of r is 0.8 which means that there is a positive correlation between the two variables. After my analysis, we can conclude that poverty does influence the number of COVID cases with all other factors held constant. So, counties with more impoverished people are more likely to have more covid cases.

The second question I'm answering is "Does more poverty mean more deaths?". To do this, I decided to compare the number of covid deaths in each county to the number of people in poverty in that county. I plotted these values and got the following result.
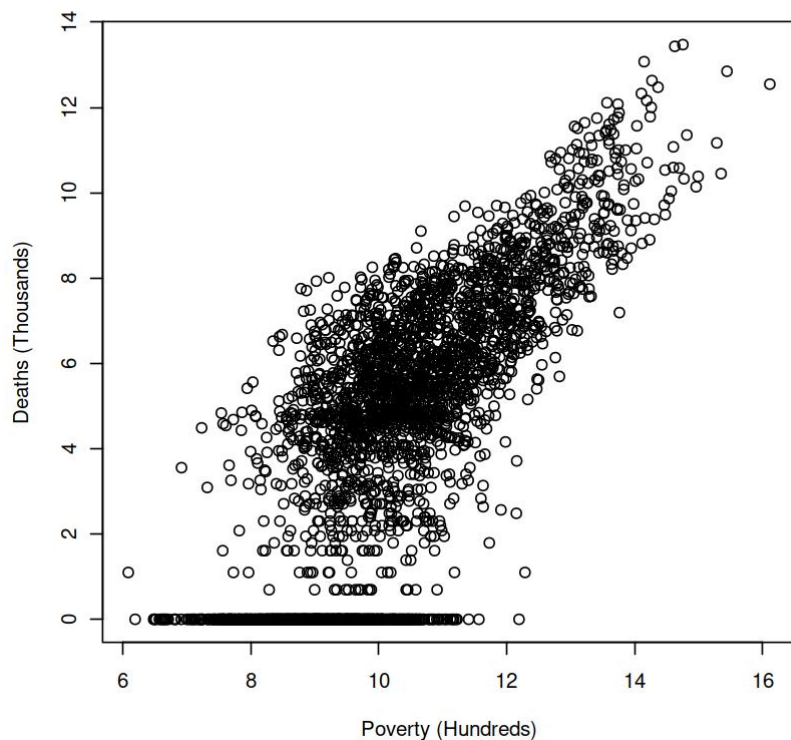
<u>Figure 4.</u>

By looking at the graph in figure 4, we can also see that there is a positive correlation between the two values. As poverty increases there is also an increase in the number of deaths for the disease. To confirm the significance of the graph, I did a Pearson's correlation test and found the p value to be 2.2×10^{-16} which shows that it is statistically significant. I also found that the value of r is 0.7 which means that there is a positive correlation between the two variables. After my analysis, we can conclude that poverty does influence the number of COVID deaths with all other factors held constant. So, counties with more impoverished people are more likely to have more covid deaths.

Conclusion

In conclusion, the analysis above confirms the following: The more impoverished people in a county the more covid cases there are and the more impoverished people there are the more covid deaths in a county. This study does confirm Patel et al (2020)'s claim that more factors must be considered when calculating the level of vulnerability of people to the virus. It also helps Peters (2020)'s argument that rural populations required different measures that large metropolitan areas. This is because more impoverished people are likely to be found in rural areas than in large metropolitan areas.

Works Cited.

Garcia, Laurindo. *Kaggle*.2020, https://www.kaggle.com/laurindogarcia.
Accessed 1 March 2022.

Patel, J A et al. "Poverty, inequality and COVID-19: the forgotten vulnerable." Public
health vol. 183 (2020): 110-111. doi:10.1016/j.puhe.2020.05.006

Peters, David. "Community Susceptibility and Resiliency to COVID-19 Across the
Rural-Urban Continuum in the US". Journal of Rural Health, vol.1, no.1, 2020, pp 1-2.