Abstract

There has been recent media attention on the phenomenon of the low-income being linked to higher COVID-19 positive test rates. This investigation is designed to determine if that was the case in Washington state. Using publicly available data sources, this study demonstrates that a test positive COVID-19 individual has a 26.88% chance of being from a low-income county of Washington state. Please note that the low-income represents a smaller population in Washington state (32.66%).

Income Impact on COVID-19 Positive Cases using Conditional Probability

Claire Gong

March 2021

Contents

1 Introduction

The chance of catching COVID-19 is influenced by multiple factors. In this study, I want to focus on the following factor: income disparities.

If the patient is positive, what is the probability that the patient comes from a county with lower income?

When deciding the scope of this study, I have chosen to limit it to a state level instead of a country level. The data drawn and the comparisons made for this study is exclusive to Washington state and its counties in order to limit the amount of external factors that would influence people's chances of catching COVID-19. Such factors include state protocols such as mask mandates or business opening capacity restrictions.

2 Formulation

Let H = high income county and C = a positive COVID-19 case. The joint probability of H and C is equal to the conditional probability of H given C multiplied by the probability of C.

$$P(C, H) = P(C|H)P(H) = P(H, C) = P(H|C)P(C) \tag{1}$$

Here, is the joint probability of C and H:

$$P(C, H) \tag{2}$$

The probability of C given H:

$$P(C|H) \tag{3}$$

The probability of H:

$$P(H) \tag{4}$$

The probability of H given C:

$$P(H|C) \tag{5}$$

The probability of C:

$$P(C) = P(C|H)P(H) + P(C|\neg H)P(\neg H) \tag{6}$$

where $\neg$ is the logical negation.

This study requires us to find the probability of a COVID-19 test positive individual from a county with a lower income, or H given C:

$$P(H|C) = \frac{P(C|H)P(H)}{P(C|H)P(H) + P(C|\neg H)P(\neg H)} \tag{7}$$

3 Data

This project requires cross-referencing of multiple online data sources.

- The positive COVID-19 cases by county data is reported by the New York Times as of March 23, 2021.
- To clarify, for the fixed income threshold, we consider low-income to be under the US median household income, in which according to the US Census Bureau, is $68,703 in 2019. Household earnings over that number will be considered as high-income.
- To determine the probability of high income households, I used Washington state's population data and the median household income of each Washington county.

4 Computation

4.1 Implementation

This project was done in two phases: $P(H|C)$ with a fixed income threshold and $P(H|C)$ with a variable income threshold.

I started with using Excel to calculate the probabilities of the fixed income threshold. Then I moved to Python for calculating variable income threshold. Please see my GitHub for the code and the Excel sheet.

4.2 Fixed Income Threshold

As stated in the data section, the fixed income threshold will be set at $68,703.

$P(H)$ The probability of a high-income household, EQ-4, was calculated by summing up the populations of the county's that had a high-income (4,985,670) and dividing that by the total Washington population (7,404,107) which resulted in $P(H) = 0.6734$.

$P(C|H)$ The probability of a testing positive for COVID-19 given the individual is from a high-income household, EQ-3, was computed by adding COVID-19 cases of counties that were considered to have an average high-income household (204,062) and dividing that with the total population of Washington state (7,404,107). $P(C|H) = 0.0276$

$P(C|\neg H)$ The probability of testing positive for COVID-19 given the individual is from a low-income household was calculated using the same method as $P(C|H)$, except for needing to meet the requirement of being from a low-income county (154,700).

$P(\neg H)$ The probability of a low-income household, was calculated by observing the fact that 1 - $P(H) = P(\neg H)$.

Therefore, by using EQ-7:

$$P(H|C) = \frac{0.0276 \times 0.6734}{0.0276 \times 0.6734 + 0.0209 \times 0.3266} = 0.7312 \tag{8}$$

4.3 Variable Income Threshold

Specifically in 5% increments from 0 to 100 percent of the maximum median household income county in Washington state ($ 89,881). See the appendix for a table (Figure-1) and graph (Figure-2) that elaborate on the relationship of the probability of coming from a low-income county given a positive COVID-19 case ($P(\neg H|C)$).

5 Conclusion

The probability of being a high-income household county given a positive COVID-19 case is 0.7312. In other words, there is a 73.12% chance a test positive COVID-19 individual is from a high-income county. This means that there is a 26.88% chance of a test positive COVID-19 individual that is from a low-income county.

6 Appendix

median household income threshold	P(not H|C)
4494	0.0000
8988	0.0000
13482	0.0000
17976	0.0000
22470	0.0000
26964	0.0000
31458	0.0000
35952	0.0000
40446	0.0000
44941	0.0001
49435	0.0140
53929	0.0626
58423	0.0892
62917	0.1161
67411	0.2334
71905	0.5299
76399	0.6397
80893	0.7452
85387	0.7452
89881	0.8795

Figure 1: Table of probability of coming from a low-income county given a positive COVID-19 case ($P(\neg H|C)$) as function of income threshold

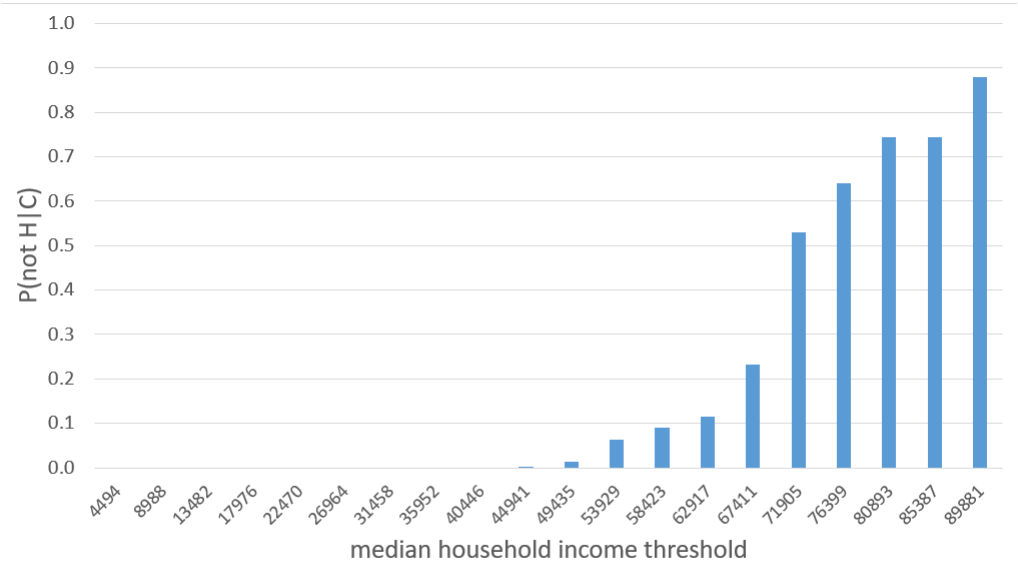

Figure 2: Chart of probability of coming from a low-income county given a positive COVID-19 case ($P(\neg H|C)$) as function of income threshold