Analysis of the Effect of Education and Health Insurance on Lifespan

Abstract

The purpose of our analysis is to investigate the effect of education and health insurance on lifespan. The Medical Expenditure Survey from the IPUMS dataset will be used to investigate these variables further. To model the effects of education and health insurance on lifespan, our analysis will use a multiple regression model. The results of this investigation was a clear association between lifespan and health insurance coverage; however, the education category did not prove to be a significant predictor for lifespan. This research adds to the already profound evidence that health insurance coverage may contribute to life expectancy.

Background and Significance

This project focuses on the relationship between lifespan, status of health insurance, and education. Evidence suggests that people with higher levels of education live longer, which may also be exacerbated with the provision of health insurance, which is also a more likely characteristic of those with higher levels of education.

Currently, health insurance is considered an extremely contentious point in politics. Unlike most countries, the US lacks a universal health care system. In its place, America is host to an assemblage of systems, both public and private, that cover many - but not all - residents. As costs for medical treatment continue to rise, unabated by the competitive pharmaceutical market, the US finds itself at the center of a decades-long debate over the necessity of a national healthcare system. The US is split along the lines of a plan for universal Medicare, and the implemented ObamaCare plan only furthered this divide. According to the US Census Bureau, in 2019, 92% of Americans reported having health insurance. While this statistic may appear all-encompassing, this means that at least 26.1 million people did not have access to health insurance coverage during 2019.

Our research aims to uncover the consequences of this steep divide between the haves and the have-nots; how a disparity of health insurance coverage affects American residents.

The primary hypothesis is that those with a higher level of attained education have a longer lifespan. The secondary hypothesis is that those with health insurance coverage will have longer lifespans.

Methods

Data

This study will use data from IPUMS-MEPS. The Medical Expenditure Survey (MEPS) set of large-scale surveys of families and individuals, their medical providers, and employers across the United States. MEPS is the most complete source of data on the cost and use of health care and health insurance coverage. MEPS provides data from the longitudinal survey of U.S. health care expenditures and utilization. We will pool data only from the year 2014 in order to make our study more succinct.

Variables

Response Variable

The response variable in our analysis will be the measure of lifespan, as in how long an individual has lived. We will model this numerical variable directly.

Explanatory Variables:

Our analysis of lifespan includes two explanatory variables: education and health insurance status.

We sorted the education variable into 5 distinct categorical levels:

> “No_HS” = Grade 8 or less
>
> “HS” = Highschool diploma, GED, or GED equivalent
>
> “SomeCollege_NoDeg” = Some college, no degree, no 4yr degree
>
> “College” = AA Degree, Bachelor’s Degree
>
> “Advanced” = Master’s Degree and higher

Health Insurance Status, represented by the HINOTCOVE variable, is a binary response variable, with a subject either having health insurance (Yes) or not (No).

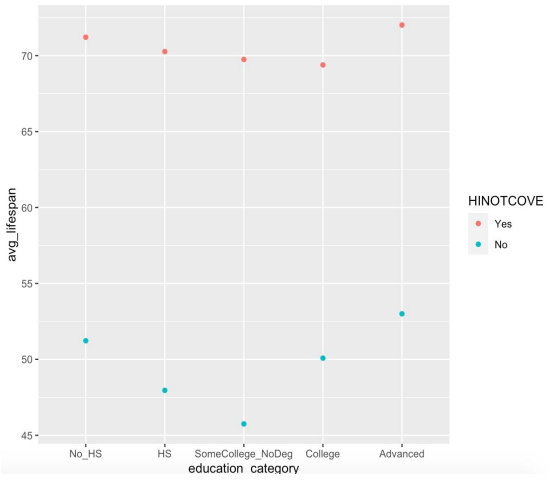
We filtered out missing values, such as the unknown education levels, other degrees, or NIU.

Analysis

We will use an interactive multiple regression model to analyze the primary and secondary hypothesis.

(Fig 1: Graph Modeling Average Interactive Effects of Education and Health Insurance on Lifespan)

(Fig 2: Regression Summary for Model)



This multiple regression interactive model highlights three things that can affect the lifespan:

1. Obtained Education Level

The first five rows of the summary table describe the effect of an individual's level of education on their lifespan.

The Intercept row shows the average lifespan of an individual who has not attained any education level, according to our education categories. Individuals who do not have a high school degree have an average lifespan of 71.2145 years.

According to our model, having a highschool degree decreases one's lifespan by an average of 0.9465 years compared to those without a degree. Therefore, on average, we can expect individuals with a highschool degree to have a lifespan of 70.7495 years.

Compared to those without a high school degree, the expected lifespan of individuals who have obtained some college education will be 1.46 units lower. Therefore, on average, we can expect individuals with some college education to have a lifespan of 69.7545 years.

Similarly, individuals with a college degree will have an average lifespan expectancy 1.83 units shorter than individuals without a highschool degree. Individuals with a college degree will have an average lifespan expectancy of 69.3845 years.

However, according to the model, those with advanced degrees beyond a college education have an increased lifespan of 0.80 years compared to those with no highschool degree. Individuals with advanced degrees will have on average, an expected lifespan of 72.0145 years.

2. HINOTCOVENo: Absence of Health Insurance

According to our model, individuals without health insurance will have a lifespan expectancy decreased by 19.98 years on average, when taking all other variables into account.

The variable HINOTOVENo (the effect of not having health insurance) decreases lifespan by 19.98 years on average. The pink points on the graph represent those that have health insurance and the blue points represent those who do not have health insurance. The average of those who do have health insurance and those who do not, 19.98 years is the difference in lifespan between them.

3. Interaction of Health Insurance and Education Level

The last 4 rows of the summary table (education category) detail the effect of the unique combination of education and health insurance.

The education_categoryHS:HINOTCOVE row shows the combined effect of having a high school degree and no health insurance decreases lifespan by 2.32 years, on average, compared to the average effects of the "HS" category and no insurance. Therefore, those who

have a combination of no health insurance and a HS degree will live an average of 47.96 years according to the model.

High-school, no health insurance calculation:

y^ = 71.2145 + (-0.9465) + (-19.9887) + (-2.3193) = 47.96 years

The education_categorySomeCollege_NoDeg:HINOTCOVENo row indicates the combined effect of having some college education and no health insurance decreases lifespan by 4.0 years, on average, compared to the average effects of the “some college education” category and no insurance. The individuals who fall into this category have a predicted lifespan of 45.75 years, on average.

Some college, no health insurance calculation:

y^ = 71.2145 + (-1.4661) + (-19.9887) + (-4.0097) = 45.75 years`

The education_categoryCollege:HINOTCOVENo row shows the combined effect of having a college degree but no health insurance actually increases life expectancy by 0.68 years compared to the average effects of the “college” category and no insurance The model shows that these individuals have a predicted lifespan of 50.08 years, on average.

College degree, no health insurance calculation:

y^ = 71.2145 + -1.8264 + -19.9887 + 0.6839 = 50.08 years

Finally, the last row, education_categoryAdvanced:HINOTCOVENo, details that the combined effect of having an advanced degree but no health insurance increases lifespan by 0.97 years compared to the average effects of the “advanced degree” category and no insurance. According to the model, those with an advanced degree but no health insurance live an average of 53.0 years.

Advanced degree, no health insurance calculation:

y^ = 71.2145 + 0.8011 + -19.9887 + 0.9731 = 53.0 years

Results

After filtering and creating the education category which contains the education level of participants, our sample size was 1149. Our filtered data also contained year of birth and death and of course health insurance status, to determine the lifespan of each individual. As stated previously, our data reveals there is an association between lifespan and health insurance coverage; however, the education category and lifespan had a very weak association. Based on our model we could see there was 19.98 years in lifespan depending on the individuals health

insurance status. Across all education levels, health insurance status matters in our model due to the small p-value (< 2.2e-16). The large p-values (p>0.05) between education categories indicate that the differences in lifespan based on education level are not significant.

Discussion

Our analysis assessed the relationship between health insurance, education level, and lifespan. Based on our results, no significant relationship was found between lifespan and education level, rejecting the primary hypothesis that a higher level of attained education results in a longer lifespan. However, there is a clear significant relationship between lifespan and whether or not an individual has health insurance. Health insurance coverage proved to be a significant predictor for lifespan. Across all education levels, the lack of health insurance significantly decreased lifespan with no statistically significant difference between each education level.

The results of this investigation show that the secondary hypothesis, which states that those with health insurance coverage will have longer lifespans, is proven to be true. Individuals who are insured live an average of 19.98 years longer according to our model.

Limitations

We suspect that demographic and socio-economic factors will be important to control for since they would be associated with both sexual orientation and with health. We will want to control variables such as race and sex, as these variables may all have an effect on lifespan, health insurance coverage, and education.

Future research directions could take into account variables such as race and sex. These variables may have a possible impact on lifespan, health insurance coverage, or education. For example, by using the variables studied in this dataset, in addition to variables that dictate race or sex, a possible relationship could be discovered.

Conclusion

It is clear from our analysis, health insurance status is significantly associated with average lifespan while education level has no significant effect on lifespan. This analysis on lifespan, health insurance, and education emphasizes the importance of health insurance coverage in regards to living a longer life, while highlighting the disparities between those with and without insurance.

References

1. IPUMS. (n.d.). Retrieved December 17, 2020, from https://ipums.org/

2. Bureau, US Census. “Health Insurance Coverage in the United States: 2019.” The United States Census Bureau, The United States Census Bureau, 15 Sept. 2020, www.census.gov/library/publications/2020/demo/p60-271.html.

Data Appendix

Checking Regression Assumptions:

Normality

```r
model <- lm(formula = Lifespan ~ education_category * HINOTCOVE, data = filtered_data)
```

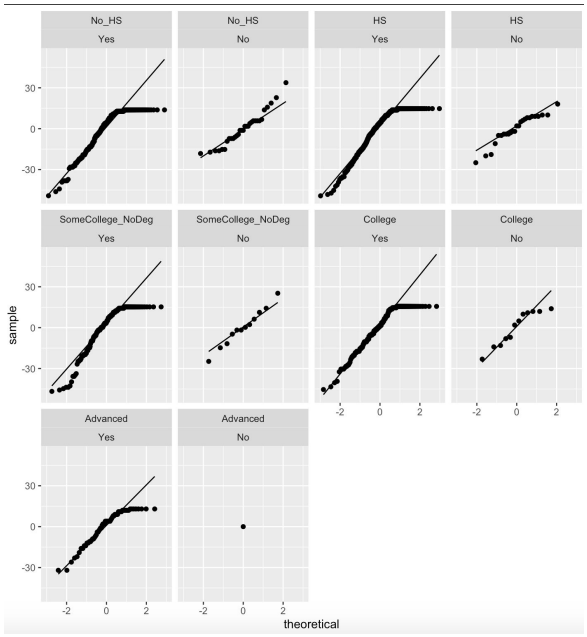
```r
filtered_data <- mutate(filtered_data, Lifespan_hat = predict(model),
                        residual = residuals(model))
```

```r
ggplot(filtered_data, aes(x = residual))+
  geom_histogram()+
  facet_wrap(~education_category*HINOTCOVE)+
  xlab("Residuals")
```

```r
ggplot(filtered_data, aes(sample = residual))+
  geom_qq()+
  geom_qq_line()+
  facet_wrap(~education_category*HINOTCOVE)
```

Histogram Plot of Residuals:

QQ Plot of Residuals:



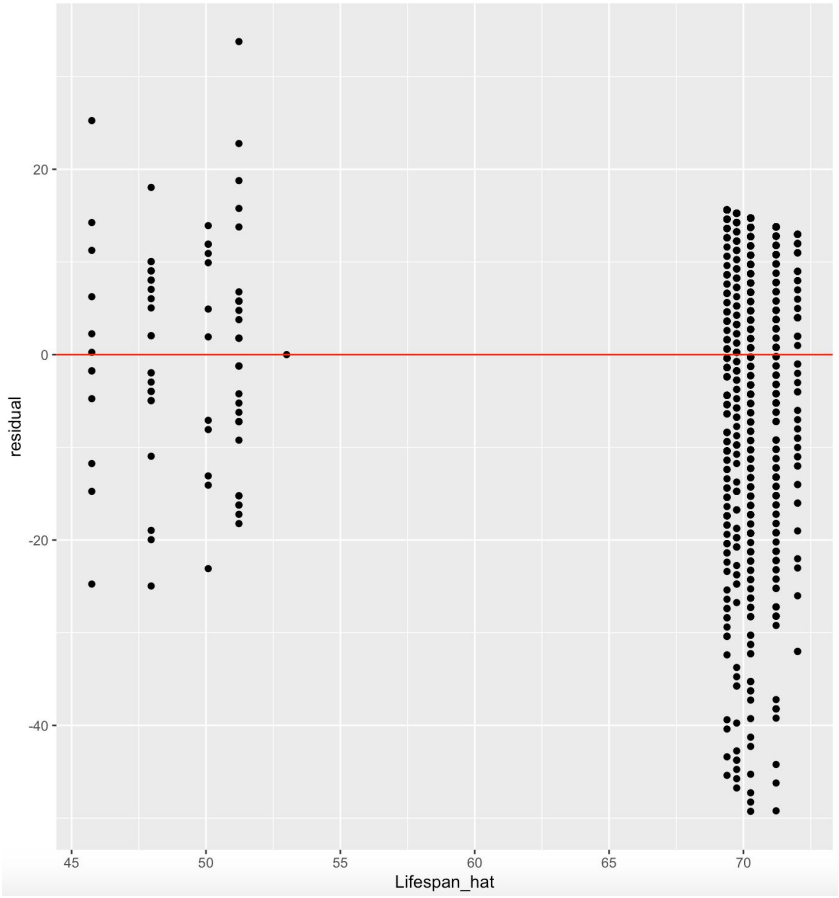
Based on these graphs, we can see that the strongest influence of normality is the HINOTCOVE(Health Insurance Status) variable. According to our data, when people do not have health insurance (HINOTCOVE:No), the data is normal. However, when people do have health insurance (HINOTCOVE), the data appears to be truncated.

We assume that this is so because those with health insurance tend to live longer. Therefore, the people with health insurance's data do not generally reach the upper bounds of their average

expected lifespan.Because people with health insurance have longer lifespans, we are unable to predict their lifespans towards the upper part of the scale. However, because people without insurance tend to have shorter lifespans, we have more regular, normal data that gives us information about the longer lifespans of people without health insurance.

Linearity and Equal Variance

Scatter Plot of Residuals versus Expected Lifespan:

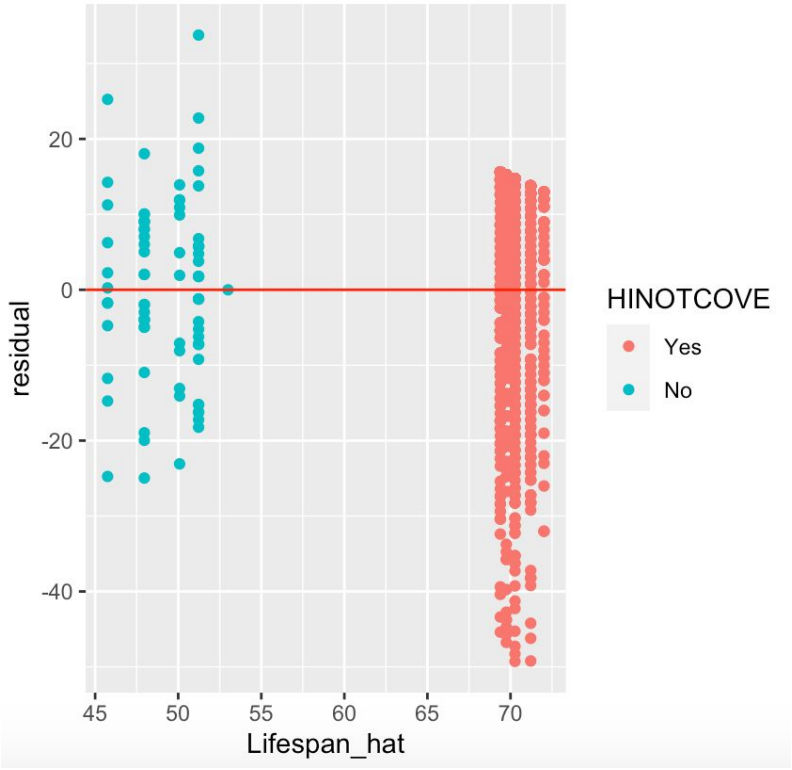


The residuals vs. width_hat plot shows us the degree of linearity and constant variance conditions in the data. It appears that the linearity assumption is slightly violated, as some of the data points fall beneath the zero line as we move from left to right across the graph.

The equal variance assumption does not appear to be met. We can see that there are more points clustered tightly together at the beginning and the end of the graph, rather than equally across the

board. We can assume that the data follows this pattern as the groups of points appear to be grouped by health insurance status, where those with health insurance tend to have distinctly higher expected lifespans than those without.

We can see that this assumption holds true in this graph:



Independence

According to our data, the independence assumption is not violated, as each person was only sampled once in the collected data. Therefore, no test subject's lifespan will influence another test subject's lifespan. Similarly, one's health insurance status or education level does not affect another's.

Code:

****Normality****

```{r}

model <- lm(formula = Lifespan ~ education_category * HINOTCOVE, data = filtered_data)

```

```{r}
filtered_data <- mutate(filtered_data, Lifespan_hat = predict(model),
                        residual = residuals(model))
```

```{r}
ggplot(filtered_data, aes(x = residual))+
  geom_histogram()+
  facet_wrap(~education_category*HINOTCOVE)+
  xlab("Residuals")
```

```{r}
ggplot(filtered_data, aes(sample = residual))+
  geom_qq()+
  geom_qq_line()+
  facet_wrap(~education_category*HINOTCOVE)
```

****Linearity and Equal Variance****

```{r}
ggplot(filtered_data, aes(x = Lifespan_hat, y = residual, color = HINOTCOVE))+
  geom_point()+
  geom_hline(yintercept = 0, color = "red")
```