Investigating the Association Between Poverty and Daily Alcohol Consumption

By: John Tsirigotis, Lingxuan Kong, Savannah Howard

Abstract

The purpose of this paper is to investigate the relationship between poverty and daily consumption of alcohol while adjusting for age, education level, gender and other factors.

Methods included conducting a literature review to form sound base and interaction models, boxplots[1] and correlation plots[2] to visualize the associations between the outcome and covariates as well as the interactions among the covariates, sequential testing and summary tables to determine which covariates were statistically significant, and then some influence and model diagnostics to validate the final model presented.

The main finding was that higher levels of poverty were associated with an increased amount of alcoholic drinks consumed per day and higher educational levels and increased age were strongly linked to lower levels of poverty. A second finding was that there were differences in each race between the education achieved and gender with respect to poverty level.

Poverty is a complicated situation that contains many components and is linked with many behaviors, habits, and activities. Being able to identify the main statistically significant elements of poverty can help us mitigate the rise of poverty levels in our society.

Introduction

There are thousands of research studies discussing the topic of poverty and its defining features. This is a highly examined topic since poverty is a universal aspect of

[1] Refer to Boxplots in the Supplemental Information
[2] Refer to the Correlation Plots in the Supplemental Information

every society and its presence is constant. One of the most prominent characteristics among the higher class is advanced educational levels and because of this, lower levels of education is associated with poverty. In 2015, UNESCO conducted a study to determine if there is a relationship between educational level and income inequality. Their findings included distinct gender disparities in which females from low-income families were more likely than males to have less educational opportunities. In addition to this, it was concluded that "low levels of education and acquisition of that education can hinder economic prosperity, which would, in turn, slow down poverty reduction" (UNESCO Institute for Statistics, 2017). Overall, it was confirmed that little education and educational connections thwart monetary growth and result in a lower income. Later in 2018, the National Center for Education Statistics did a study to look at the differences in educational levels and opportunities between different races. Their findings were that between 2000-2016, high school completion rates increased across all races. However, the continued gap between other races and Whites have remained the same with Whites having higher high school completion rates. Educational retention for Whites and Asians are higher than other races as they are more likely to attain higher degrees and, ultimately, achieve higher median incomes (de Brey, 2019).

The disparities among races are not only seen in the education system, but are also prominent in varying poverty levels. An investigation was done in the 2000s to look into how race, income level, and health status were all linked together. First, they looked into the variation among socioeconomic statuses and where the majority of each race falls. It was determined that Asians and Whites had higher overall wealth and socioeconomic status compared to their Black and Hispanic counterparts, with Blacks and Hispanics being

more likely to be destitute (Williams, 2016). With many researchers exploring the tie between race and poverty level, it was discovered that gender and age within each race has some association with income level. The contrasts between genders within each race and how that relates to their economic status were looked into. It was found that "black... women experience extreme economic hardship owing to being both women and members of a minority group. As compared with whites, however, gender inequality among other minority groups is relatively small" (Elmelech and Lu, 2004). This illustrates that there are some gaps in income-level between genders, with males generally having a higher income than females, and the magnitude of these gaps differs among races.

Studies have been done that look at the relationship between age and poverty level. According to data produced by the 24/7 Wall St., "American children are far more likely to live below the poverty line than adults" (Stebbins and Frohlich, 2020). The data showed that people aged 18 years and younger had the highest poverty rate within their age group. Plus, Erin Duffin facilitated a study that looked into how age and gender together associated with income levels. It was established that age levels 24 years and younger had the highest poverty rates and those poverty rates decreased with older age groups. Within each age group, females always had a higher proportion of their population living in poverty than males (Duffin, 2020).

It was inspected whether this trend holds while looking at just the relationship between gender and income level without any interaction with age. In this examination, it was shown that females make 82 cents for every dollar that males make, among full-time year round workers (Semega, 2019). It's worth noting that, over the years, the wage gap between males and females has decreased, but is still present. Still, females having a lower

income than males contributes to the reason why females have higher poverty rates than males.

Two other features among the differing economic classes are alcohol consumption and abuse. In a study that was conducted from data collected between 2000 and 2005, there is a linkage between economic class and their alcohol consumption. It was found that "poverty ratios were associated with alcohol use and problems, whereas overall income inequality was not" (Karriker-Jaffe, 2013). Therefore, this study will carry out statistical analyses that will allow us to really observe the association between alcohol consumption and poverty levels.

In this study, we will construct a model that will take into consideration all of these covariates that have been proven to be associated with poverty from other investigations. Also, interactions between covariates will be included in the model to better our understanding on how all these predictors relate to our poverty outcome. Ultimately, we will begin to design a model that will be composed of covariates that represent the multiple defining characteristics found within underprivileged groups.

Methods

To investigate whether there is an association between the level of poverty a person experiences and alcohol consumption, we utilized the R dataset NHANES. The NHANES dataset was collected by the NCHS, who utilized a series of health and nutrition surveys since the 1960s. The target population for NHANES is "the non-institutionalized civilian resident population of the United States". The data that was used to conduct our research includes a sample size of 10,000 people that were resampled from another version of NHANES (NHANESraw in R), where racial minorities were oversampled. We created a

dataset for our own study that prevents oversampling of specific groups of people, removed all NAs from the *Poverty* and *AlcoholDay* variables defined in the NHANES dataset, and only looked at people aged 20-69 years old which resulted in a sample size of 4125 observations to statistically analyze.

The main predictor of our model is the *AlcoholDay* variable which is defined as the "average number of drinks consumed on days that the subject drank alcoholic beverages". Upon reviewing the data, variables were selected based on research from literature reviews described in the introduction, that would need to be adjusted for when conducting a linear regression of alcohol level onto poverty score. These variables were age (*Age*), race (*Race1*), education (*Education*), and gender (*female—Gender was dichotomized into this variable such that a female subject would be assigned a value of 1 and male a value of 0*).

The chosen outcome variable for the model was *Poverty,* which is defined as a ratio of family income to poverty guidelines; therefore, the smaller the poverty score, the more poverty experienced. It was discovered that there are many elements that delineate poverty and most of them are influenced by each other because poverty is such a complicated subject.

As stated above, our main predictor of poverty (*Poverty*) is the amount of alcohol consumed on days where drinking occurs (*AlcoholDay*), while accounting for the confounders of age (*Age*), race (*Race1*), education (*Education*), and gender (*female*). From the literature review, it was found that there are statistically significant interactions between age and race (*Age*Race1*), age and gender (*Age*female*), race and educational

level (*Race1*Education*), and race and gender (*Race1*female*).[3] This resulted in the main effect model:

$$Poverty_i = \beta_0 + \beta_1 alc_i + \beta_2 Age_i + \beta_{3-6} Race_i + \beta_{7-10} Education_i + \beta_{11} female_i + {}_i$$

and the final interaction model:

$$Poverty_i = \beta_0 + \beta_1 alc_i + \beta_2 Age_i + \beta_{3-6} Race1_i + \beta_{7-10} Education_i + \beta_{11} female_i + \beta_{12-15}(Age * Race1)_i +$$

$$\beta_{16-31}(Race1 * Education)_i + \beta_{32-35}(Race1 * female)_i + {}_i$$

To start, we corrected for outliers in the *AlcoholDay* variable by categorizing it into a new variable, *alc_level*. The new variable was created by taking the median and standard deviations of *AlcoholDay* (2 and 3.23637, respectively), and creating four levels based on these values. The first value was the median plus one-half a standard deviation (since a full standard deviation seemed too high), the next category was one full standard deviation more (sd + 3/2 medians), and so on.

Next, we want to figure out whether our samples can represent the whole dataset or we must constrain our conclusion to a certain group. To compare continuous variables, we carried out a two sample t-test, and to compare categorical variables, we applied a chi-square test.[4] Among these six variables, all of the continuous variables have significant differences in means of cases included and cases dropped, while only two categorical variables have significant differences in their compositions. This result means that the conclusions derived from our statistical analyses can only be applied to groups that meet certain conditions. To solidify literature review inferences involving the interaction terms, several boxplots and correlation plots were created to visualize findings. The *alc_level*

[3] Refer to Table 1 in Supplemental Information for more detailed descriptions of each covariate
[4] Refer to Table 2 in Supplemental Information

variable was regressed on *Poverty* while adjusting for all covariates, and the result was statistically significant ($p<0.001$). The interaction terms were included in the new model via sequential testing and a sensitivity test was run against the base model. The conclusion of this sensitivity test was that the outcome variable, *Poverty*, was sensitive to the interaction variables ($p<0.001$), so they were included in the analysis. The steps of analysis included a scatterplot of *alc_level* vs. *Poverty*, checking LINE assumptions, using robust standard errors, and finally some model diagnostics such as Cook's distance and a studentized residuals plot.

Results

First, a scatterplot was conducted between the *alc_level* and *Poverty* variables.[5] It is unsurprising that the scatterplot has four vertical lines because we have a categorical variable as the x-axis (*alc_level*), and a continuous variable as the y-axis (*Poverty*). Also, a regression line was fitted on this graph to see the correlation between *alc_level* and *Poverty*. Although the R-squared value is relatively low (~5.4%), suggesting that *Poverty* is only slightly explained by *alc_level*, it does imply that there is a negative correlation between our main predictor and outcome, which intuitively makes some sense (i.e. the more someone drinks alcohol on a daily basis, the more poverty they are likely to experience). Next, we looked at the summary table statistics of the main effect model[6] and the final interaction model.[7]

In our main effect model, we included five variables, two of which are continuous variables and the rest categorical. Black males with an 8th grade education have a starting

[5] Refer to *alc_level* vs. *Poverty* scatterplot in Supplemental Information
[6] Refer to Main Effect Model Summary in Supplemental Information
[7] Refer to Final Interaction Model Summary in Supplemental Information

poverty score around 0.4 and every unit increase of 1 year of age is associated with an average increase of 0.025 in poverty score, slowly decreasing poverty. Our model summary table shows that, with the other covariates adjusted for, an increase in alcoholic drinks consumed per day is associated with a decrease in poverty score by 0.164 points, on average. This result may be caused by money spent on buying alcohol and less productivity.

While adjusting for the other covariates, people in other races have a better economic condition than Hispanics and Mexicans, who have slightly higher poverty scores, while White people have nearly a 0.6 point higher poverty score on average, which means their poverty score doubled compared to Black people when both are young. This result could be due to there being more educational opportunities available or being in a more privileged situation that enables those races to more easily further their education.

An increase of one level of education for a subject that has completed 8th grade is associated with a large increase in poverty score, all else being equal. From high school to college, every level is associated with approximately a 0.5 increase in poverty score, and graduate education is associated with an average of 0.8 points higher poverty score than college education. On average, poverty score also differs between males and females. Females, on average, have a 0.12 lower poverty score than males.

To further investigate whether the covariates we included have effects on each other, we added interaction terms into our model. For some coefficients like age, its effect on poverty score differs little between the races and genders. When including interaction variables, the conclusion that drinking more alcohol is associated with a lower poverty score still stands, all else being equal. For other races, with the other coefficients adjusted, if they only have an 8th grade education level, they have lower poverty scores compared to

Black people. However, they benefit most from 9-11th grade education and interestingly, other races also benefit most from some college, college education, and graduate education. We can also conclude that the effect of gender on each race is different. For example, for Black people the difference between male and female poverty score is the largest, and in the other races, the conclusion is reversed in which females have higher poverty scores compared to males while adjusting for other covariates. Based on this information, we carried on the rest of the influence and model diagnostics with our final interaction model.

In the residual model, we can conclude that the linearity assumption is met. However, the variation may not be necessarily constant since it's hard to tell based on the graph, thus our model potentially might not satisfy the equal variance assumption.[8] To fix this, we fit more "correct" heteroskedastic standard errors by using robust standard errors of coefficients. When we plot our new robust and heteroskedastic standard errors of estimated coefficients, we get a scatterplot that looks more like the type of plot we are looking for to satisfy the equal variance assumption.[9] Furthermore, because the points on this plot are now random since no obvious pattern is seen, we can reasonably conclude that our independence assumption is also met. Next, we can tell that the normality assumption holds roughly true in the case of our regression model because of the Normal Q-Q plot.[10] This plot illustrates the ordered standardized residuals (the y-axis), and the theoretical residuals if our residuals are truly normally distributed (x-axis). If our model is normally distributed, x would equal y, thus displaying a straight diagonal line, which our standardized residuals roughly do in this graph.

[8] Refer to Fitted Values vs. Residuals Plot in Supplemental Information
[9] Refer to Estimates vs. Std Errors Plot in Supplemental Information
[10] Refer to Normal Q-Q Plot in Supplemental Information

Now that the assumptions for our regression model have been met, we can diagnose any strongly influential points that are throwing our model off. From this Cook's Distance Plot, we see three observations that are strongly influential on our model.[11] Although our most influential subject in our data frame did not meet the threshold (0.001), upon removing it from the data it did not strongly affect the linear model. Next we looked at the studentized residuals plot to detect outliers with absolute-value threshold of 3.[12] After removing the most influential data line from our data frame, the regression did not suggest that the outlier strongly affected the association between *alc_level* and *Poverty*. Therefore, the decision to keep all observations was made.

Conclusions and Discussion

Our conclusive results are restricted to people aged 20-69, and based on cases with full records of *Poverty*, *AlcoholDay*, *Gender*, and *Race1* (NAs were taken out from the dataset). From both models, it was demonstrated that an increase of drinking alcoholic beverages per day will decrease poverty score (i.e. worse socioeconomic conditions), while adjusting for race, gender, age, and education. It was also presented that there are differences within each race when it comes to educational achievement and gender with respect to poverty level. Finally, people with higher education and older age tend to have a higher poverty score. This implies that the connection between education or age and poverty score may be directly linked. Poverty as an outcome may have a more complex interaction with education or age as people with high poverty scores may be able to support themselves enough to be able to obtain that higher education or established jobs which ultimately increases their income, and therefore makes their poverty score larger.

[11] Refer to Cook's Distance Plot in Supplemental Information
[12] Refer to Studentized Residuals Plot in Supplemental Information

Works Cited

de Brey, Cristobal, et al. "Status and Trends in the Education of Racial and Ethnic Groups in 2018." *U.S. Department of Education*, Feb. 2019.

Duffin, Erin. "Poverty Rate in the U.S. by Age and Gender 2019." *Statista*, 21 Sept. 2020, www.statista.com/statistics/233154/us-poverty-rate-by-gender/.

Elmelech, Yuval and Lu, Hsien-Hen. "Race, ethnicity, and the gender poverty gap." *Social Science Research* vol. 33, 1 (2004): 158-182. doi: 10.1016/S0049-089X(03)00044-9.

Karriker-Jaffe, Katherine J et al. "Income inequality, alcohol use, and alcohol-related problems." *American journal of public health* vol. 103,4 (2013): 649-56. doi:10.2105/AJPH.2012.300882

Semega, Jessica. "Payday, Poverty, and Women." *The United States Census Bureau*, U.S. Census Bureau, 10 Sept. 2019, www.census.gov/library/stories/2019/09/payday-poverty-and-women.html.

Stebbins, Samuel, and Thomas C. Frohlich. "The Poverty Rates for Every Group in the US: From Age and Sex to Citizenship Status." *USA Today*, Gannett Satellite Information Network, 28 Feb. 2020, www.usatoday.com/story/money/2019/11/06/united-states-poverty-rate-for-every-group/40546247/.

UNESCO Institute for Statistics. "UNESCO Study Reveals Correlation between Poverty and Education." *UNAA*, UNESCO, 27 June 2017,

www.unaa.org.au/2017/06/28/unesco-study-reveals-correlation-between-poverty-and-education/.

Williams, David R et al. “Understanding associations among race, socioeconomic status, and health: Patterns and prospects.” *Health psychology : official journal of the Division of Health Psychology, American Psychological Association* vol. 35,4 (2016): 407-11. doi:10.1037/hea0000242

Group Project Contributions

Lingxuan:

- did boxplots and correlation plots for methods section
- wrote part of methods/results/conclusion sections
- did summary tables
- did skeleton of powerpoint

John:

- did all the R coding
- graphs for results (model and influence diagnostic sections)
- wrote part of methods/results
- proofread/edited paper

Savannah:

- Abstract/introduction
- wrote part of methods/results/conclusions
- Bibliography
- put together supplementary document