ECON 320: Econometrics

Ted Jesus C. Chua

September 3, 2019

Introduction

Many studies have examined the consequences—and potential benefits—of alcohol consumption at the individual and societal level. The merits of this socially accepted behavior arguably lies in moderation, the frequency of which there is no universal standard. Alcohol is very much prevalent in many cultures on a global scale, albeit public policies and tax disincentives aimed in reducing consumption. Such initiatives are perhaps for a valid reason: the Center for Disease Control and Prevention (CDC) reports that approximately 88,000 deaths in the United States (U.S.) between 2006 and 2010 were caused by "excessive alcohol use."[1] But given that over 12,000,000 people in the U.S. died during this five-year period, the aforementioned figure is not as alarming.

Alcohol abuse is nonetheless associated with a poor quality of life and negative economic costs, including healthcare expenditures and alcohol-related crimes. In terms of productivity, a study found statistically significant losses in earnings (not higher unemployment) among alcoholic males; the losses were much greater among those who started drinking before 15 years old.[2] This cohort of individuals who began alcohol use at an earlier age than the other group are certainly different in other characteristics, namely exogenous factors relating to family upbringing.

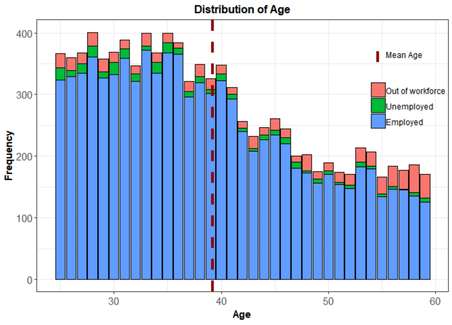
In that regard, I am ultimately interested in distinguishing the effects of genetic and environmental factors on alcoholism. Within the scope of my project, I want to explore this domain with family history and stress indicators. The survey data did not ask respondents about stress levels, such that I aim to utilize job prospects and family size as proxies for stress in my classification model. Indeed, a family may certainly serve as a support system for alcoholism; and even encourage individuals to seek employment. A potential interaction between these two predictors will be considered.

Data

```
alcohol<-alcohol
needToBeFactor<-c(1:2,6,8:20,24:28)

for (i in needToBeFactor){
    alcohol[,i]<-as.factor(alcohol[,i])
}

alcohol$Area<-NA
for (i in 1:nrow(alcohol)){
        if (alcohol[i,13]==1){
        alcohol[i,]$Area<-"Northeast"
    }
    else if (alcohol[i,14]==1){
        alcohol[i,]$Area<-"Midwest"
    }
    else if (alcohol[i,15]==1){
        alcohol[i,]$Area<-"South"
    }
    else {
        alcohol[i,]$Area<-"Other"
    }
}

alcohol$Health<-NA
for (i in 1:nrow(alcohol)){
        if (alcohol[i,9]==1){
        alcohol[i,]$Health<-"Excellent"
    }
    else if (alcohol[i,10]==1){
        alcohol[i,]$Health<-"Very Good"
    }
    else if (alcohol[i,11]==1){
        alcohol[i,]$Health<-"Good"
    }
    else if (alcohol[i,12]==1){
        alcohol[i,]$Health<-"Fair"
    }
    else{
        alcohol[i,]$Health<-"Poor"
    }
}

alcohol$Area<-as.factor(alcohol$Area)
alcohol$Health<-as.factor(alcohol$Health)
alcohol$Alcoholic_Parent<-as.factor(ifelse(alcohol$mothalc==1 | alcohol$fatha
alcohol<-alcohol[,c(1:2,28,3:7,34:36,21,23)]
```

Alcohol Abuse?	n	Median Age*	% with Alcoholic Parent(s)*	% Employed*
No	8848	38	0.17	0.90
Yes	974	38	0.26	0.87

Health	n	Median Age	% of Alcoholics	% Employed	Mean Family Size
Excellent	4085	36	0.08	0.94	2.69
Very Good	2965	37	0.10	0.93	2.81
Good	2016	39	0.11	0.88	2.78
Fair	525	45	0.12	0.71	2.73
Poor	231	48	0.12	0.30	2.48



Pearson's Chi-squared test

data: . X-squared = 7.8577, df = 2, p-value = 0.01967

There are significant differences in the prevalence of alcoholism based on employment status ($p < 0.05$).

Empirical Model: Logistic Regression