Measuring the Effect of Mortality Rate and Alcohol Consumption Using Minimum Legal Drinking Age - Carpenter and Dobkin

David Pham - 1005349053

June 21, 2021

Abstract

This report seeks to follow the work of Christopher Carpenter and Carlos Dobkin, extending their study of alcohol consumption and mortality rate in the United States from 1997-2004. We concentrate on different regression discontinuity methods, such as a linear approach and another quadratic approach, and focus on the relationship of mortality rate as a function of age, with the minimum legal drinking age (MLDA) as the threshold/cutoff. It was found that both linear and quadratic approaches showed statistically significant increases in average mortality rate after the threshold (from 8-10%). These results show that alcohol does play a role in mortality rate, and more can be done to regulate drinking and show its risks.

Introduction

Alcohol consumption is a worldwide phenomenon that plays an important social role in many cultures. There are many instances where a drink or two is not a concern, but excessive consumption can be dangerous. It is very easy to have a dependency or even become addicted to alcohol, and its capacity to affect our cognitive abilities make it lethal, especially for young adults.

A study in 2009 conducted by Christopher Carpenter and Carlos Dobkin studied the effect of alcohol consumption and mortality rate by comparing deaths from 1997-2004 before and after the minimum drinking age in the United States [1]. They used regression discontinuity and other methods to conduct their analyses using survey data on Americans and their alcohol consumption habits, as well as historical mortality information. What they found was that there was a statistically significant increase in mortality rate after the minimum drinking age, and zoomed in on motor vehicle accidents, suicides and alcohol-related deaths as the primary causes [2].

In this report, I would like to replicate and extend Carpenter and Dobkin's work by further examining the historical mortality information, and add a STA304 twist to it. More specifically, I will only look at the relationship between mortality rate and age by fitting complex regression discontinuity models. The control group will be individuals below the threshold/minimum drinking age, and the treatment will be those above it.

Lastly, the overall main objective is to determine whether or not the minimum drinking age in the United States is appropriate, or if it should be modified. Currently, the minimum legal drinking age (MLDA) is 21. The results of this paper can also be extended worldwide, as many countries (including Canada) have a much lower MLDA. I expect that with using my models, the results should match those of Carpenter's and Dobkin's; however, I believe there will be a sharp jump at the threshold and it will decline again.

Data

Data Collection Process

The data from the original study were gathered from the National Center for Health Statistics (NCHS) on confidential mortality information from American citizens during the period of 1997-2004 [3]. Luckily, all the observational data were readily available for free on the American Economic Association's website. We oblige to the terms of service, agreeing that we only plan to use the data for personal usage. All the data files were .sas files, so we had to convert them to file types that R can process. This took quite a while to figure out, but thankfully only one file was needed for this report. After removing some unnecessary variables from the dataset, we can now proceed. In summary, we have a stratified random sample (by state) of the US population from 1997-2004, containing 52 samples of aggregate groups [4].

A drawback of this data is that it is very outdated. Seeing that this information was from 20 years ago, it is possible that the results/analyses have changed throughout the years. In fact, one of the peer feedback suggested that I try and find more recent data in order to improve relevancy; however, it was too difficult to find data that had confidential information as such. Furthermore, since this was a probability sampling technique, the sample will be representative of the total population. However, this can be subject to "non-participation" by randomly selected deceased individuals. This can present moral issues respective to friends and family of the sampled members, but this is in a grey zone.

Data Summary

In total, here are all the variables we will be looking at in this report:

- `agecell`: The aggregated average age of the bin of individuals upon death.
- `all`: The average mortality rate per 100,000 people in the United States, aggregated from 1997-2004. This variable is the sum of internal and external deaths.
- `internal`: The average mortality rate per 100,000 of individuals in the United States who passed away due to medical conditions and/or internal organ failures, aggregated from 1997-2004.
- `external`: The average mortality rate per 100,000 of individuals in the United States who passed away due to external factors, aggregated from 1997-2004. The five variables below are subsets of this variable and make up the sum of this variable as well.
- `alcohol`: The average mortality rate per 100,000 of individuals in the United States who passed away due to alcohol poisoning, aggregated from 1997-2004.
- `homicide`: The average mortality rate per 100,000 of individuals in the United States who passed away due to homicide, aggregated from 1997-2004.
- `suicide`: The average mortality rate per 100,000 of individuals in the United States who passed away due to suicide, aggregated from 1997-2004.
- `mva`: The average mortality rate per 100,000 of individuals in the United States who passed away due to motor vehicle accidents, aggregated from 1997-2004.
- `externalother`: The average mortality rate per 100,000 of individuals in the United States who passed away due to other external reasons, aggregated from 1997-2004 [5].

Note that a deceased individual is able to have more than one cause of death assigned to their death certificate.

Since most of the data wrangling/cleaning was done outside of R, the process was generally straight-forward. After converting the dataset from .sas to a .csv file, I had removed all the columns with fitted values since I was going to create all the analyses and graphs from scratch. Furthermore, I create a `threshold` indicator variable, where the cell value is equal to 1 if the age bin is 21 or over, 0 otherwise. We are all done with parsing the data, and can look at some important numerical summaries.

To begin, we will take a look at some plots.

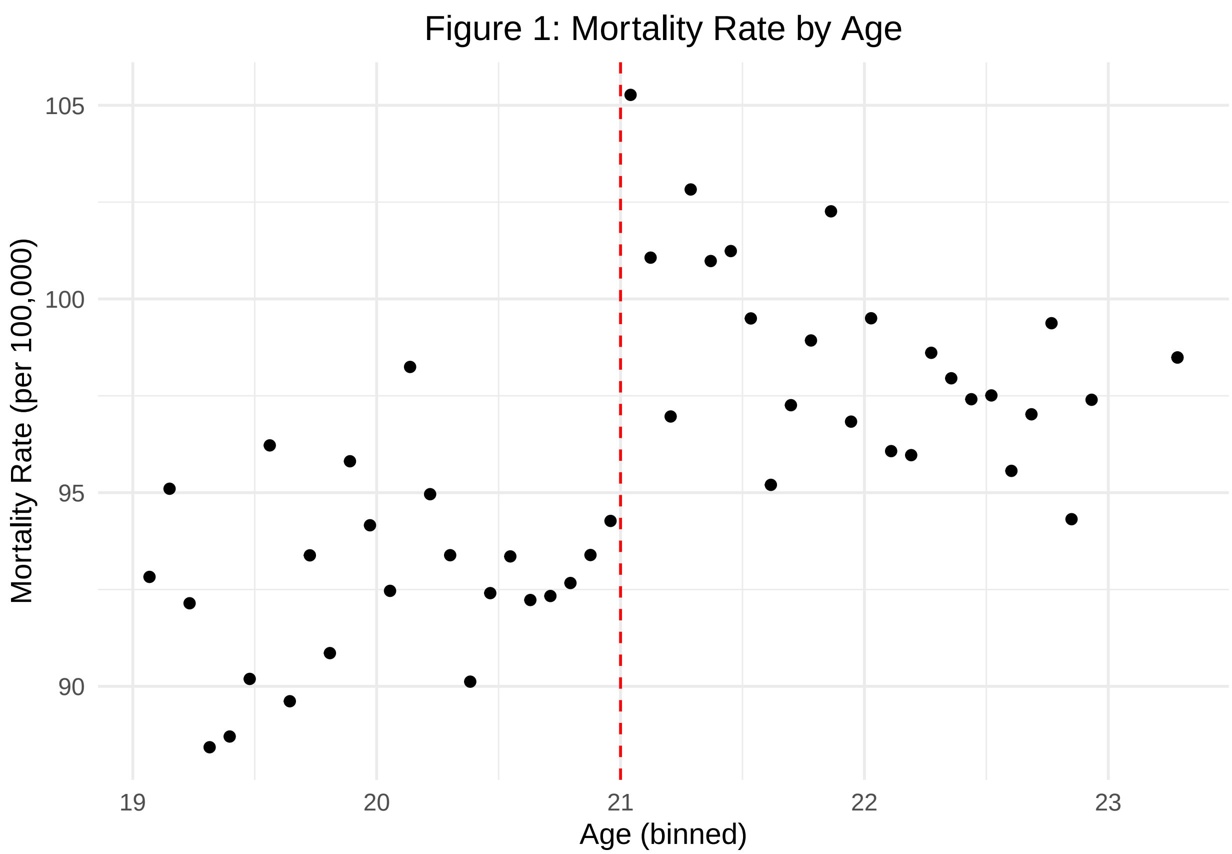

Figure 1 depicts a simple plot of the aggregate mortality rate of the United States from 1997-2004, by age [6]. The red dashed line in the middle represents the minimum drinking age, placed at age 21. On opposite sides of the line, we initially see that this regression can be modelled as a piecewise function, as there is a noticeable jump at the threshold.

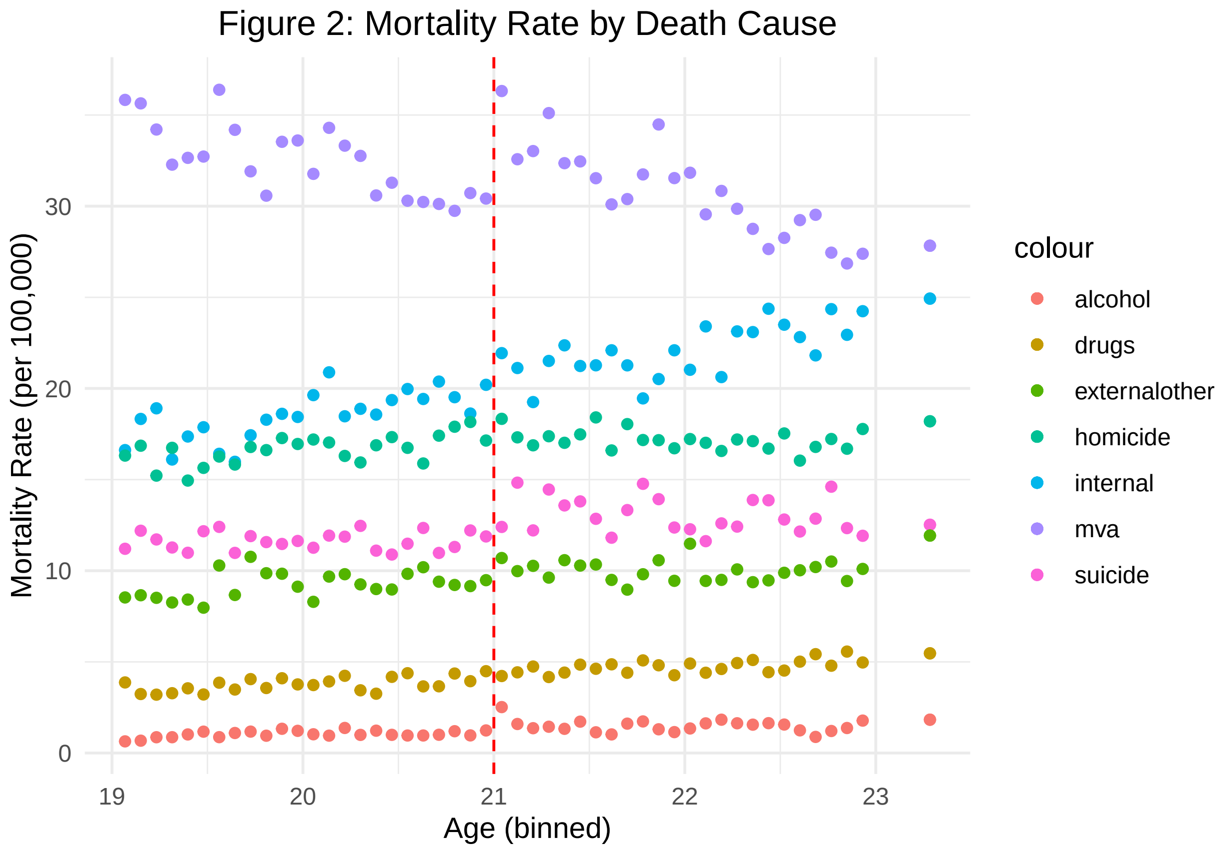

Figure 2 is a plot of mortality rate as a function of age, but we specifically examine by death cause [7]. This is similar to Figure 4 in Carpenter and Dobkin's study [8]. There are some minor discontinuities in each category, but the most noticeable jump is the mortality rate increase due to motor vehicle accidents after age 21. Following our hypothesis, this makes sense as vehicle accidents are highly correlated with alcohol consumption.

Table 1: Means for all variables in the dataset, by threshold group. [9]

Variable Name	Mean for Non-Treatment Group	Mean for Treatment Group
Mortality Rate Average	92.803	98.539
Internal Average	18.511	22.28
External Average	74.292	76.305
Alcohol Average	1.032	1.509
Drugs Average	3.768	4.788
Average of Other External Causes	9.216	10.131
Homicide Average	16.641	17.261
MVA Average	32.462	30.557
Suicide Average	11.634	13.028

The table above shows all the aggregate means for mortality rate by death cause, grouped by the age threshold. To recap, an age bin has a threshold value of 1 if it is 21 or over, and 0 otherwise. Across the board, we see that the mean mortality rates for age groups 21 and over are all higher than the younger groups, with the surprising exception for motor vehicle accidents despite the noticeable jump on the graph. This should give us enough motivation to use regression discontinuity for the report and see if there is plausible evidence for our hypothesis.

All analysis for this report was programmed using `R version 4.0.4`.

Methods

Next, we will be discussing the methods that will be used in the report. We are primarily going to focus on the general mortality rate as a function of age (i.e, Figure 1) and use regression discontinuity, as well as statistical inference to see if there is a significant difference in death rates before and after the minimum drinking age. Two types of regression models will be used: a flexible linear model and a quadratic model.

Assumptions

Firstly, following Cunningham's assumption, we assume that the threshold/cut-off we are using is "known, precise, and free of manipulation" [10]. We are strictly focusing on the minimum drinking age as the threshold, where it is impossible to manipulate age upon death. Hence, regression discontinuity will be valid where the no treatment group is below the threshold, and the treatment group is above it.

Furthermore, the last key assumption is that the forcing function is continuous [11]. This is obvious since the forcing variable is based on legislation, and there are sufficient observations below and above the cutoff.

Equations

The linear equation for the model looks like this:

$$all = \beta_0 + \beta_1 * agecell + \beta_2 * I(agecell >= 21) + \beta_3 * [agecell * I(agecell >= 21)] + \epsilon$$

where

- `all` is the response variable of interest (i.e, the true mortality rate we are trying to estimate)
- β_0 is the intercept term for the average mortality rate for the treatment group (at the discontinuity).
- $\beta_1 * agecell$ is the variable that accounts for the relationship between age bin and mortality rate. To interpret this, β_1 represents the change for every one unit increase in age bin.
- $\beta_2 * I(agecell >= 21)$ represents the indicator/dummy term that separates the treatment and control groups. We will be focusing on this variable, as it will show us the difference in mortality between both groups. **An important thing to note is that this variable is represented by the `threshold` variable we had created earlier.**
- $\beta_3 * [agecell * I(agecell >= 21)]$ is the interaction term between age and the cutoff of the minimum drinking age [12].
- ϵ is the error term.

Finally, the quadratic equaion for the model:

$$all = \beta_0 + \beta_1 * agecell + \beta_2 * agecell^2 + \beta_3 * I(agecell >= 21) + \beta_4 * [agecell * I(agecell >= 21)] + \beta_5 * [agecell^2 * I(agecell >= 21)] + \epsilon$$

where this time, we have introduced a squared term for the age bin variable, as well as another interaction term between the minimum drinking age threshold and age squared. All other variables and coefficients remain the same.

Results

For our models, we specify linear models and regress on the mortality rate on the indicator variable, as well as the respondents' ages by centering them around the threshold value of 21 years using the `I()` function [13].

Firstly, the plot with linear slopes and its associated summary table:

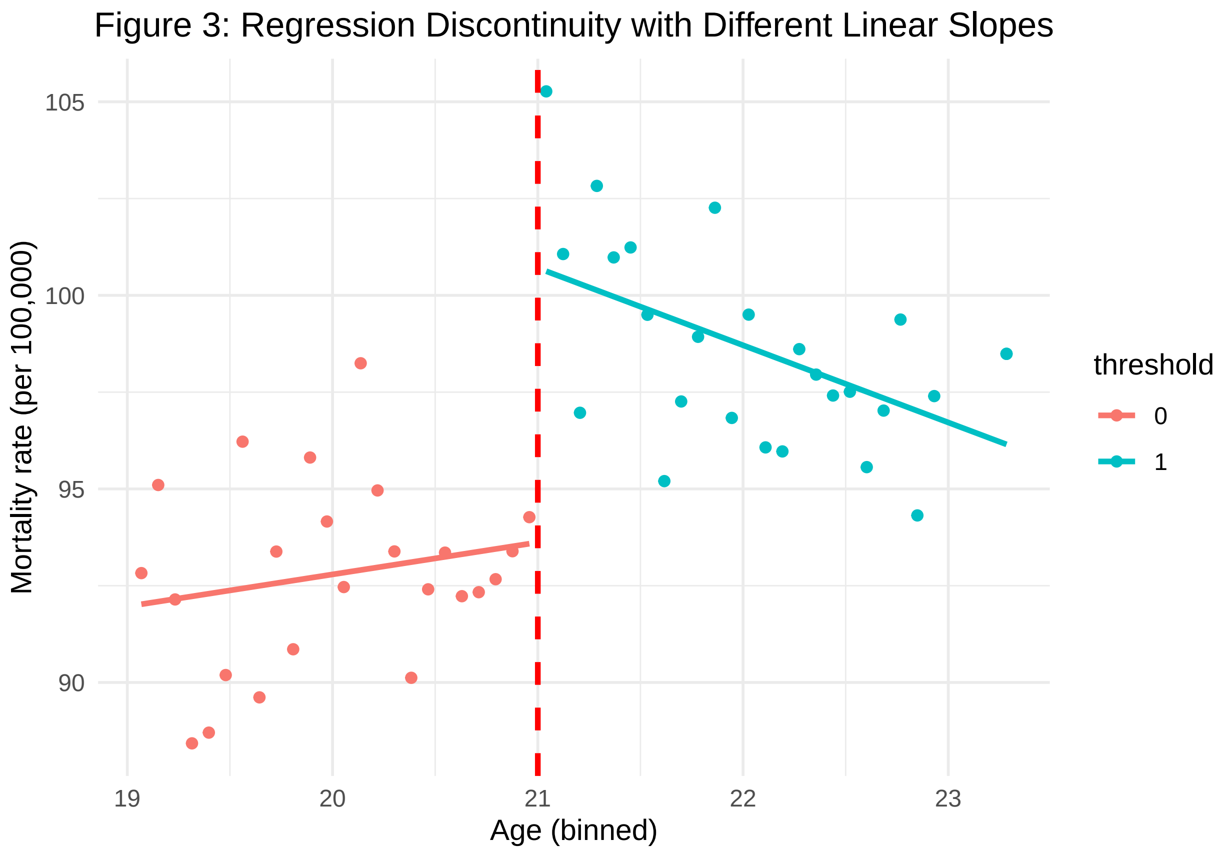

Table 2: Summary for Linear Model. [14]

Variable Name	Estimate	P-Value
Intercept	93.618	< 2e-16
Age (bin)	0.827	0.324
Indicator Term	7.087	1.8e-06
Interaction Term	-2.822	0.013

We see that the slopes before and after the cutoff differ greatly. Right after the discontinuity, the intercept jumps up and the trend line is larger than the entire slope line before the cutoff, despite being negative. Furthermore, if we observe the table, the intercept, dummy variable, and interaction are all statistically significant as they are below the benchmark significance level of 0.05. Firstly, this means that it is very plausible for the intercept term of the control group to be around 93 deaths per 100,000 people. Next, the threshold being significant implies that the mortality rate per 100,000 for individuals reaching the minimum drinking age is higher by 7.087, on average. Finally, the interaction term is slightly significant, which means that with caution, we can observe that being above the threshold, as well as your age, play a role in mortality rate.

With a bit more persistance and effort, we can do better than this. Lastly, we shall take a look at the quadratic approach.

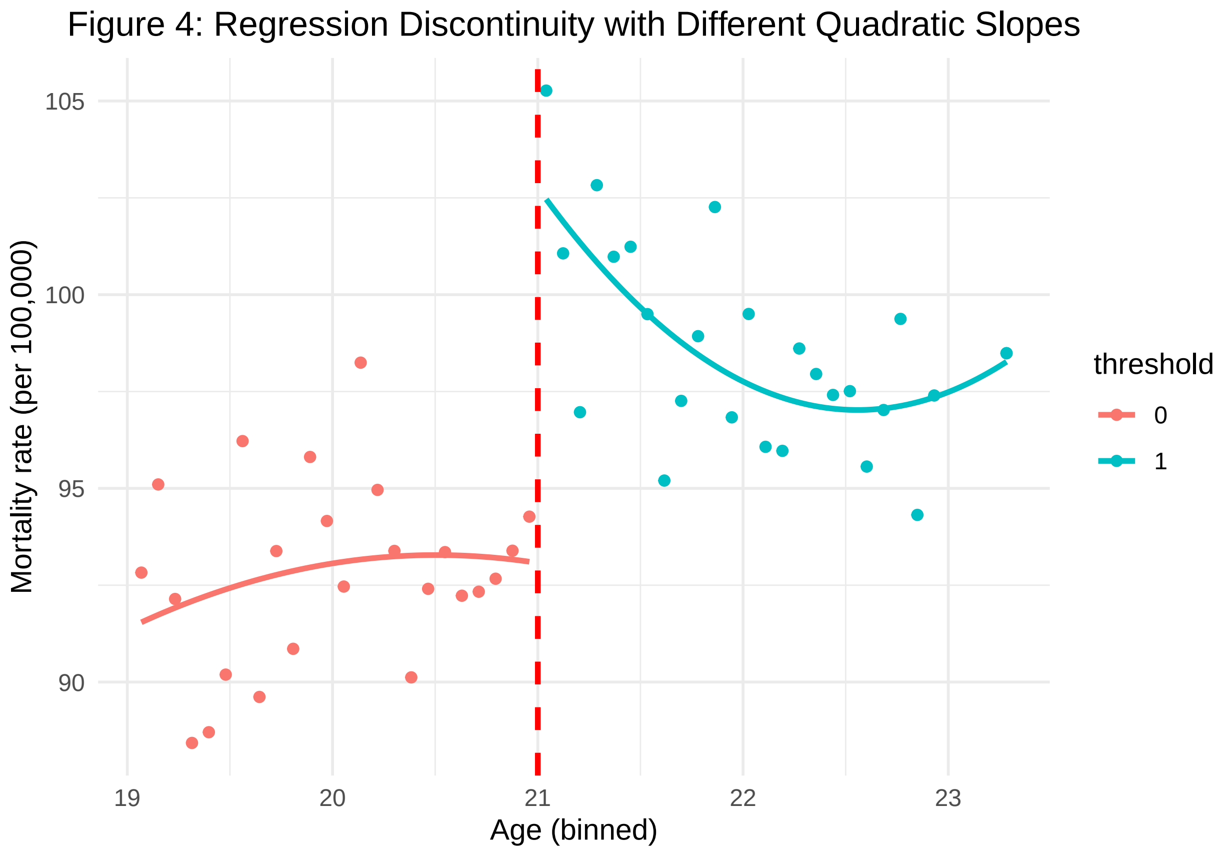

Table 3: Summary for Quadratic Model. [15]

Variable Name	Estimate	P-Value
Intercept	93.073	< 2e-16
Age (bin)	-0.831	0.797
Indicator Term	9.688	4.9e-06
Interaction Term	-6.535	0.116
Age Squared	-0.84	0.597
Interaction Term with Squared Age	3.204	0.099

In this final plot, we see that the regression lines (curves in this case!) fit the data much more accurately than the linear model; in fact, the adjusted R-squared value is also higher [16]. Furthermore, as predicted in the original hypothesis, the regression curve after the threshold starts up very high, but slowly dips back down and plateaus. This time, the indicator term, our `threshold` variable, has increased to 9.688; hence, with this model, the average for mortality rate per 100,000 individuals after reaching the minimum drinking age is now 9.688 higher than those below the threshold, compared to 7.087 in the previous model.

Conclusions

In summary, we tried to replicate and extend Carpenter's and Dobkin's work and see if the minimum legal drinking age was appropriate at the time in the United States. We used linear and quadratic regression discontinuity models to measure the average change in mortality rate before and after the MLDA threshold. Furthermore, we examined each death cause briefly to look at the ones most correlated with alcohol consumption.

In Figure 2, we can match Carpenter's and Dobkin's observations and see that motor vehicle accidents and suicides are the most correlated with alcohol consumption, leading to higher mortality rates after the MLDA.

Furthermore, when looking at our regression discontinuity methods, we saw that plotting different linear slopes between the threshold led to a 7.087 average increase in mortality rate (per 100,000 individuals). When using the quadratic approach, we had a **9.688** average increase in mortality rate after the MLDA, which was even higher than the linear model. Therefore, using the quadratic model, the mortality rate of individuals after the threshold is 10% higher on average, which is extremely close to the results found in original report [17].

Our hypotheses and results were spot-on and very close to Carpenter's and Dobkin's, despite using different regression discontinuity methods. A 10% average increase in mortality rate seems like a very sharp jump, so it is possible that more attention needs to be paid to the minimum legal drinking age.

Weaknesses

Some obvious weaknesses are that the data is outdated, and replicating a study may have led to bias. A study from over a decade ago using data from over 20 years ago may not fully address or be representative of the population today. Although both reports differ from each other in terms of methods, I still used their data and hence some bias can be observed.

Furthermore, the sample size was extremely small and may have lacked key information about the population that we could not have known about.

Next Steps

Moving forward, it would be nice if more updated data was readily available (with preferably a larger sample). I was looking around for historical mortality data from the late 2010's in the United States and Canada, but it was unfortunately not what I was looking for. This way, we can strengthen our past findings and advocate more on the dangers of alcohol consumption.

Discussion

In all, we can agree that alcohol consumption can be dangerous if it is not regulated properly. Focusing in on the youth, young adults just want to have a good time and make the most out of their prime age with friends. However, we can sometimes fail to neglect the consequences due to cognitive deficiency after consuming alcohol. The majority of countries have their MLDA at 18, so 21 is surprisingly high for the United States [18]. If legislation believes that this is an appropriate age, more needs to be done in order to expose the risks of alcohol consumption. Increased funding in public service announcements, enhanced security and other factors can contribute to saving young lives.

Bibliography

1. Carpenter, Christopher, and Carlos Dobkin. 2009. *The Effect of Alcohol Consumption on Mortality: Regression Discontinuity Evidence from the Minimum Drinking Age.* American Economic Journal: Applied Economics 1 (1): 164–82. https://doi.org/10.1257/app.1.1.164.
2. Carpenter, Christopher, and Carlos Dobkin. 2009. *The Effect of Alcohol Consumption on Mortality: Regression Discontinuity Evidence from the Minimum Drinking Age.* American Economic Journal: Applied Economics 1 (1): 164–82. https://doi.org/10.1257/app.1.1.164.
3. Carpenter, Christopher, and Carlos Dobkin. 2009. *The Effect of Alcohol Consumption on Mortality: Regression Discontinuity Evidence from the Minimum Drinking Age.* American Economic Journal: Applied Economics 1 (1): 164–82. https://doi.org/10.1257/app.1.1.164.
4. Carpenter, Christopher, and Carlos Dobkin. 2009. *The Effect of Alcohol Consumption on Mortality: Regression Discontinuity Evidence from the Minimum Drinking Age.* American Economic Journal: Applied Economics 1 (1): 164–82. https://doi.org/10.1257/app.1.1.164.
5. Grolemund, G. (2014, July 16) *Introduction to R Markdown.* RStudio. https://rmarkdown.rstudio.com/articles_intro.html.
6. Wickham et al., (2019). *Welcome to the tidyverse.* Journal of Open Source Software, 4(43), 1686, https://doi.org/10.21105/joss.01686
7. Wickham et al., (2019). *Welcome to the tidyverse.* Journal of Open Source Software, 4(43), 1686, https://doi.org/10.21105/joss.01686
8. Carpenter, Christopher, and Carlos Dobkin. 2009. *The Effect of Alcohol Consumption on Mortality: Regression Discontinuity Evidence from the Minimum Drinking Age.* American Economic Journal: Applied Economics 1 (1): 164–82. https://doi.org/10.1257/app.1.1.164.
9. Allaire, J.J., et. el. *References: Introduction to R Markdown.* RStudio. https://rmarkdown.rstudio.com/docs/.
10. Cunningham, Scott, *Causal Inference: The Mixtape.* http://www.scunning.com/causalinference_norap.pdf
11. Wu, Changbao and Mary E. Thompson, (2020), *Sampling Theory and Practice*, Springer.
12. Carilli, Tony. *R Companion to Real Econometrics.* https://bookdown.org/carillitony/bailey/
13. Carilli, Tony. *R Companion to Real Econometrics.* https://bookdown.org/carillitony/bailey/
14. Allaire, J.J., et. el. *References: Introduction to R Markdown.* RStudio. https://rmarkdown.rstudio.com/docs/.
15. Allaire, J.J., et. el. *References: Introduction to R Markdown.* RStudio. https://rmarkdown.rstudio.com/docs/.
16. Sheather, J. Simon, (2009), *A Modern Approach to Regression with R*, Springer.
17. Carpenter, Christopher, and Carlos Dobkin. 2009. *The Effect of Alcohol Consumption on Mortality: Regression Discontinuity Evidence from the Minimum Drinking Age.* American Economic Journal: Applied Economics 1 (1): 164–82. https://doi.org/10.1257/app.1.1.164.
18. Carpenter, Christopher, and Carlos Dobkin. 2009. *The Effect of Alcohol Consumption on Mortality: Regression Discontinuity Evidence from the Minimum Drinking Age.* American Economic Journal: Applied Economics 1 (1): 164–82. https://doi.org/10.1257/app.1.1.164.

Appendix

Supplementary Data

```
## Rows: 50
## Columns: 11
## $ agecell       <dbl> 19.1, 19.2, 19.2, 19.3, 19.4, 19.5, 19.6, 19.6, 19.7,...
## $ all           <dbl> 92.8, 95.1, 92.1, 88.4, 88.7, 90.2, 96.2, 89.6, 93.4,...
## $ internal      <dbl> 16.6, 18.3, 18.9, 16.1, 17.4, 17.9, 16.4, 16.0, 17.4,...
## $ external      <dbl> 76.2, 76.8, 73.2, 72.3, 71.3, 72.3, 79.8, 73.6, 75.9,...
## $ alcohol       <dbl> 0.639, 0.677, 0.866, 0.867, 1.019, 1.171, 0.870, 1.09...
## $ homicide      <dbl> 16.3, 16.9, 15.2, 16.7, 14.9, 15.6, 16.3, 15.8, 16.8,...
## $ suicide       <dbl> 11.2, 12.2, 11.7, 11.3, 11.0, 12.2, 12.4, 11.0, 11.9,...
## $ mva           <dbl> 35.8, 35.6, 34.2, 32.3, 32.7, 32.7, 36.4, 34.2, 31.9,...
## $ drugs         <dbl> 3.87, 3.24, 3.20, 3.28, 3.55, 3.21, 3.86, 3.48, 4.06,...
## $ externalother <dbl> 8.53, 8.66, 8.51, 8.26, 8.42, 7.97, 10.29, 8.67, 10.7...
## $ threshold     <fct> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,...
```

Supplementary Credits

https://people.ucsc.edu/~cdobkin/Papers/2009%20The%20Effect%20of%20Alcohol%20Consumption%20on%20Mortality%20Regression%20Discontinuity%20Evidence%20from%20the%20Minimum%20Drinking%20Age.pdf

There are no supplementary plots and methods. Here is a link to the original paper, where my report was inspired from. Most of the methods and analyses used in the paper are beyond the scope of this course.