Alena McLucas

Professor Dobkin

Econ 104

March 19th, 2014

Third Draft Changes

Abstract

* TA said wasn’t worth mentioning birthday effect here, so deleted it from two places to stay focused on MLDA.
* Added sentence about assumptions.

Introduction

* Added three questions of the paper from the prompt to increase clarity.
* Added three sentences in the first paragraph to frame motivation for research question.
* I deleted my last paragraph and broke down the end into two sections: regression discontinuity methods and results, and IV methods and results.
* Rewrote most of the methods for regression discontinuity.

Data

* Found National Vital Statistics System and researched to include information about the mortality data. Added citation in Appendix.

Method

* Added paragraph about regression discontinuity and why it’s an appropriate method at the beginning.
* Rearranged the subsections to fit lecture note order.
* Rewrote beginning and ending of the first paragraph in the “Figures” section to explain the importance of Figures for RDD.
* Added two paragraphs to the beginning of the “Methods”, “Regression” section that explain the parametric approach, role of the birthday effect, and how to select the best regression from figures and tables. Kept the paragraphs about how the data was prepared to keep clear.
* Rewrote entire “Instrumental Variable Estimate” section to explain what it is, why it’s important, and how it is theoretically derived.
* Wrote an “Assumptions” subsection per the request of the TA that defines each assumption. Added citation from article.

Results

* Deleted sloppy lines in “NHIS Data” and explained why β2 is equal to the change as opposed to doing all the math for over 21 line subtract by under 21 line. More clean and concise.
* Simplified same process as noted above for “Mortality Data.”
* Took off percentage signs for mortality data results, as they are not percentages.
* Moved “IV Estimates” to this section per request of TA.
* Rewrote majority of “IV Estimates” to strengthen interpretation.
* Added a paragraph for both data set results analyzing the birthday effect.

Conclusion

* Contextualized why the assumptions are important for the interpretation of the IV estimate.
* Added another paragraph that connects the questions to results, and concludes that the MLDA should not be lowered.

Appendix

* Renamed graphs as Figures and tables as Tables per the request of the TA.
* Though the TA wanted me to get rid of the graphs with each regression side by side, but they help me compare and justify my decision for the best fitting regression so I kept them in.

Alena McLucas

Professor Dobkin

Econ 104

March 19th, 2014

The Impact of the MLDA on Drinking, All Deaths, Motor

Vehicle Accidents, and Alcohol Poisoning

Abstract

The Minimum Legal Drinking Age (MLDA) has been a debated subject for decades, as the United States drinking age is three years higher than in many other countries. This paper explores the impact of the United States’ MLDA on drinking and mortality using data from the NHIS and NVSS. After identifying the best-fit regression from tests of different polynomial orders, we calculate the instrumental variable effect. The results of calculating this effect tell us that lowering the MLDA would increase mortality by about 47 deaths per 100,000 people. Assumptions to validate the use of regression discontinuity and MLDA as the instrument variable are mostly met.

Introduction

How much does the MLDA reduce the proportion of the population that drinks? How much does the MLDA reduce mortality? Do the IV assumptions validate or invalidate the use of the IV estimate? The important points covered by this paper all accumulate to calculating the amounts that the MLDA reduces the population that drinks and reduces mortality. This is pertinent because the question is frequently posed as to if we should lower or keep the MLDA. We use several methods on the data to ensure accuracy so that a definitive assertion can be made. We use 21 years of age as an instrument to determine how the MLDA affects mortality. In order to calculate the instrumental variable estimate, assumptions are made that must be examined. These results are important because they support the notion that the MLDA should not be lowered to 18.

The government regularly collects data to monitor the health of the United States, and the NHIS is one of the most polls of information. Calculations derived from these surveys are used to make statistics that are nationally applicable. There is has been no information given about how the mortality data was collected.

The method we used to interpret the jump in data at 21 years of age is regression discontinuity, which is comprised of two main components: regression tables and figures. Drinking alcohol, all deaths, motor vehicle accidents, and alcohol-related deaths data was used in linear, quadratic, and cubic regressions. We ran second regressions with a birthday variable on each of these regressions to trace the MLDA’s affects. The figures were made by trial-and-error methods, plugging in different values for bin size, x-axis and y-axis. We find that the MLDA reduces drinking by about 9.55%, and the birthday effect is not significant. The change in all deaths, motor vehicle accidents, and alcohol poisoning are 9.55, 4.35, and 0.44, respectively. The birthday affect is significant for alcohol poisoning.

The IV estimate was calculated based on the best regressions from drinking alcohol and all deaths, both quadratic. It is about 47.35 with a standard error of 10.08. After obtaining the IV estimate, we test and find that almost all of the assumptions have been met.

Data

The data about who drinks, how much, and general demographic information was collected by the National Health Interview Sample Adults Files between 1997 and 2007. The National Health Interview Survey (NHIS) is an organization that is executed through the U.S. Census Bureau to assess the welfare of people in a variety of topics relating to the “amount, distribution, and effects of illness and disability in the United States” (“About the…”). The NHIS resurveys people every 10-15 years, hence the information is most relevant to the aforementioned 10 year gap.

The mortality data came from the National Vital Statistics System (NVSS). This data is shared directly out of the Centers for Disease Control and Prevention. It collects its data through its contract with vital registration systems that record significant events like “births, deaths, marriages, divorces, and fetal deaths” (“National Vital…”). An additional note we recall from class is that this data assumes everyone’s birthday is on the 15th of the month they were born in. This biases the data downward, but only slightly.

Methods

Regression Discontinuity (RD) Design makes comparisons by leveraging changes in treatment exposure probability. It allows the slope of a line to change at the threshold, where treatment probability changes, of the forcing variable age. MLDA is a prime example, as the government tells people that they can or cannot drink based on age and will choose their treatment status. The jump before and after 21 for percentage of people that drink and number of people that die is immediately noticeable when one looks at the plotted data, which is ideal for regression discontinuity. The two components to this approach are: (1) Figures, and (2) Regressions. We will also calculate the Instrumental Variable estimate to find the true treatment effect, and examine the assumptions under which it is a valid measure.

Figures

Figures are important because they help the data tell a story by focusing on the pertinent portions and excluding irrelevant ones. In order to find the right bin size, x-axis range and y-axis range we use trial-and error as well as past experience to select the best values. We must first pick an appropriate bin size, which is the number of days that will be averaged to create one data point in the graph. We choose 32 as it strikes the balance between too noisy (from higher bin sizes) and too smooth (from lower bin sizes). The x-axis should focus on 21 years of age, or the threshold, so the range selected goes from 19 to 23. This gives us just enough room to see the trend, but not so much room that we see the sharp incline from 18 to 19 or the steady, slight decline that persists after turning 21 until 30 years of age. Lastly, the y-axis is automatically selected by R to fit the data best. We coded them anyways for clarity’s sake, making them 0.45 to 0.75 for the NHIS data, 85 to 110 for all deaths, 25 to 40 for motor vehicle accidents, and 0.0 to 3.0 for alcohol poisoning. A line from each regression found below was constructed for a visual comparison of each polynomial order to find the line of best fit for the data, meaning on average it summarizes the data the best.

Regression

To run the regressions, we are taking a parametric approach by experimenting with different polynomial orders and running them on a fixed range of 19 to 23 years of age. The higher polynomial orders in x give more flexibility in fitting the data, are less likely to be biased, but have higher variance estimates. As the data has significant slopes, we will not be running zero-order polynomials; instead, we will run linear, quadratic, and cubic regressions. The regressions are run a second time adjusting for covariates, in this case the birthday effect. By including this variable in the regression, we check that the model fits the data well and that there are no significant changes unexplained by the regression. In this case, the most likely effect that intersects with the MLDA is that it is a person’s birthday, and they are likely to celebrate that regardless of the MLDA. To assess the birthday effect’s beta, one must find its t-statistic to see if it is statistically significant. If it is, it must be contextualized in the application to see if it is economically significant.

To make the table, we construct a column in the data that contains a 1 if the person drinks and a 0 if they do not, so a subset had to be created to make sure those were the only values. This column contains binary results, so in order to average the number of people who identify with each at each age, the data had to be aggregated. The data from the NHIS spans from age 18 to 30, and so a subset of the original data was made that contained all data points between 19 and 23.

To prepare the data for the regressions, we created a dummy variable so that two lines could be derived from the same regression to represent under 21 and over 21 people. This variable is 0 if the individual is under 21, and 1 if the individual is 21 or over. This fosters a break at 21 years of age so a second line can be plotted with different characteristics. We then constructed new columns to represent the variables that would have to be multiplied together, including the interaction terms, so they wouldn’t cause errors in calculation. Lastly, a variable for the birthday effect was created that requires a person to be exactly 21 to be equal to 1 and 0 otherwise. This is testing to see if the birthday effect is unique to the 21st birthday because it is at the MLDA, or if it any jump is due to unaccounted for factors like general celebration of a birthday.

The equations for the linear, quadratic, and cubic regressions are as follows, with ‘drinks alcohol’ being replaced by ‘all deaths’, ‘motor vehicle accidents’, and ‘alcohol poisoning’ for the mortality regressions:

Linear: drinks alcohol = β­­0 + β1from21 + β2over21 + β3from21\*over21 + u

Quadratic: drinks alcohol = β­­0 + β1from21 + β2over21 + β3from21\*over21 + β4from21\*from21 + β5from21\*from21\*over21 + u

Cubic: drinks alcohol = β­­0 + β1from21 + β2over21 + β3from21\*over21 + β4from21\*from21 + β5from21\*from21\*over21 + β­6age\*from21\*from21 + β7from21\*from21\*from21\*over21 + u.

For comparison’s sake, we made a regression table that contains linear, quadratic, and cubic regression covariates with and without the birthday affect. The t-statistics for the age covariates must be calculated to determine the line of best fit from regression’s standpoint, and the t-statistic for the birthday covariate must be calculated to determine if the celebration of a birthday affects the regressions ability to capture the data’s trends well. These tables allow us to numerically look at t-statistics to determine the best one, as anything at or above 2.00 is generally acceptable. The results must be compared both in the table and visually in the figure that contains the three plotted regressions to select the best-fitting polynomial order. The jump, or birthday effect at the MLDA, can then be calculated by subtracting the under 21 line from the over 21 line.

Instrumental Variable Estimate

The instrumental variable (IV) estimate is helpful to answer the questions posed in this report because it is the true treatment affect of treatment exposure on outcome. To calculate it, the best regressions for drinking alcohol and all deaths onto age must be picked, and their coefficients and standard errors noted. The First Stage (D = B0 + B1Z + u) must be calculated, which is the effect of assignment to treatment on the treatment status. Then, we calculate the Reduced Form (Y = P0 + P1Z + u), or the effect of treatment assignment on the outcome. The coefficients from the treatment assignment variable from both derivations are then taken, and the one from the Reduced Form (P1) is divided by the one from the First Stage (B1). This yields the IV estimate, the true treatment effect.

Assumptions

Each of the following assumptions must be defined and subsequently assessed in the context of this study to see if it has been violated:

Assumption 1 (SUTVA): outcomes for each individual are not affect by treatment status of other individuals.

Assumption 2 (Ignorable Assignment): assignment of treatment status is random.

Assumption 3 (Exclusion Restriction): there is a single causal change from treatment assignment to treatment exposure to the outcome.

Assumption 4 (Nonzero Average Causal Effect of Z on D): if one averages the effect of treatment exposure on treatment status, it does not equal zero.

Assumption 5 (Monotonicity Assumption): we can assume that there would be a change for each individual with and without the treatment, though we can only observe one treatment status (Angrist 446-447).

The assumptions are evaluated in the Conclusion.

Results

NHIS Data

Table 1 contains the graphs with each regression. Visually, it would appear that the linear regression does captures the trends fairly well, but the quadratic regression summarizes the data a bit better with its notable curve for the over 21 line. The regression table with estimates of changes in drinking are in Figure 1. It has the linear, quadratic, and cubic regressions with and without the birthday effect. Looking at these results, it would appear that quadratic regression forms the lines of best fit as overall it has the highest t-statistics, meaning they are less likely to have happened due to chance. We therefore prefer the quadratic regression. Figure 2 is the final figure with the quadratic regression of drinking alcohol onto age.

The MLDA reduces drinking, and to calculate that reduction we must subtract the line representing over 21 in our quadratic regression from the line representing under 21 in the same regression. If one subtracts the under 21 regression line from the over 21 regression line with any polynomial order, they will get the value β2. This determines if the MLDA reduces drinking and by how much. For the NHIS data, we get 9.55. Therefore, the MLDA reduces drinking by about 9.55%.

The t-statistics for the birthday effects are 0.67, 1.33 and 1.20. This means the birthday effect is not significant on the 21st birthday for the percentage of the population that drinks.

Mortality Data

The graph with estimates of change in all deaths, motor vehicle accidents, and alcohol poisoning are depicted in Figures 3, 5, and 7, respectively, and the regression tables are depicted in Tables 2, 3, and 4. We prefer the quadratic regression for all deaths because, though the linear t-statistic is higher, the quadratic regression graph clearly has a defined curvature. We prefer the linear regression for motor vehicle accidents because it has the highest t-statistic and the linear line summarizes the trends of the data well. Lastly, alcohol poisoning could really be linear or quadratic, but because the t-statistic is about 2.0 higher for the linear regression we prefer it. Figure 4,6, and 8 contain the final figures with the scatterplots and preferred regressions overlaid. Figure 9 contains a graph for motor vehicle accidents, alcohol poisoning and age to put the difference in deaths into perspective.

By the same logic in the previous section, we can calculate the reduction in mortality:

All deaths (quadratic regression) = 9.55

Motor vehicle accidents (linear regression) = 4.35

Alcohol poisoning (linear regression) = 0.44

Therefore, the change in all deaths, motor vehicle accidents, and alcohol poisoning are a 9.55, 4.35, and 0.44 decrease, respectively.

The t-statistics for the birthday effects are below 2.00 for all deaths and motor vehicle accidents, meaning the birthday effect is not significant. However, the t-statistics for alcohol poisoning are 4.17, 4.15, and 2.9. This means that the birthday effect is significant, which makes sense because it is the first day a person can legally drink.

IV Estimate

The important part of this paper is proved by the IV estimate because it will tell us the additional deaths per 100,000 people at the MLDA for all deaths. As we have chosen quadratic regressions for both the NHIS data and all deaths, those are the regressions that must be plugged into the equations and code.

Plugging in the quadratic equations, 47.35 is the IV estimate with a standard error of 10.08. Dividing them, we get 47.35 / 10.08 = 4.70. A t-statistic this high means that it is unlikely these results occurred by chance and are fairly accurate. This means that, for every 100,000 people, about 47 less people will die on their 21st birthday. However, we must take caution because this IV estimate was calculated with all deaths. This would include deaths that occurred for other reasons but were included in the data set (i.e. Homicide, Suicide, Drugs, Other External Causes, etc). 47 less deaths are about the true treatment effect of the MLDA on mortality per 100,000 people.

Conclusion

In order to check if our IV method was justified, we must test the assumptions to see if they hold. The SUTVA assumption was violated in this study, as people who are drunk can cause the mortality of those that are not drunk. The ignorable assignment assumption is not violated because assignment is base purely on age. The exclusion restriction assumption is not violated because actions taken that affect mortality cannot be changed, and the assignment has a direct effect on treatment exposure which has a direct effect on outcome. The nonzero average causal effect of assignment on action is not violated because there is a causal effect of assignment on exposure. Lastly, the monotonicity assumption is violated because people who are below 21 do drink, and people above 21 do not drink by choice or for health reasons. This means that our overall results should only be taken into consideration under these pretenses and contexts.

How much does the MLDA reduce the proportion of the population that drinks? We find that the MLDA reduces drinking by about 9.55%, and the birthday effect is not significant. How much does the MLDA reduce mortality? The decrease in all deaths, motor vehicle accidents, and alcohol poisoning are 9.55, 4.35, and 0.44, respectively. The birthday affect is significant for alcohol poisoning. What is the IV estimate, and do the IV assumptions validate or invalidate the use of the IV estimate? The IV estimate is about 47, meaning that for every 100,000 people, about 47 less people will die on their 21st birthday with the MLDA at 21. After obtaining the IV estimate, we test and find that almost all of the assumptions have been met. From these results, we can conclude that the drinking age should not be lowered from 21 in the United States.

Appendix

Figures

Figure 1: Three Regressions of Drinking Alcohol on Age

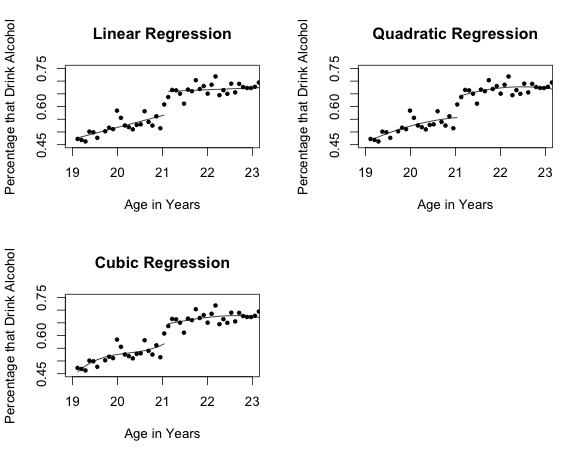


Figure 2: Quadratic Regression of Drinks Alcohol on Age



Figure 3: Regressions of All Deaths onto Age

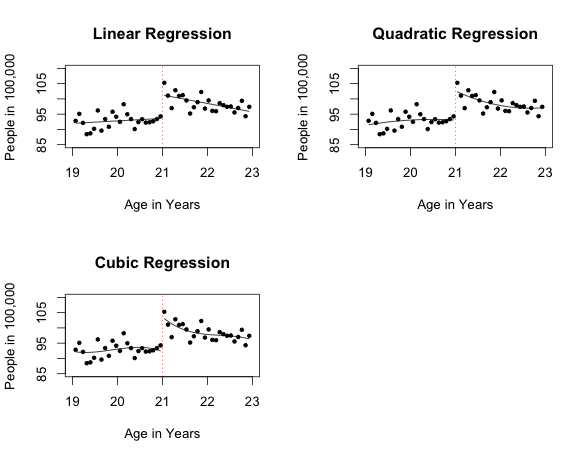


Figure 4: Quadratic Regression of All Deaths onto Age



Figure 5: Regressions of Motor Vehicle Accidents onto Age



Figure 6: Linear Regression of Motor Vehicle Accidents onto Age



Figure 7: Regressions of Alcohol Poisoning onto Age

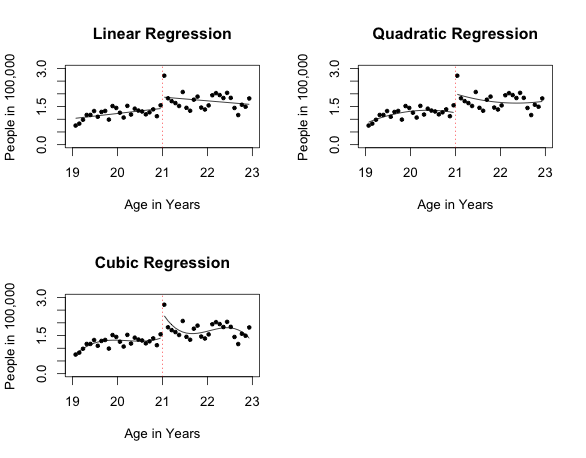


Figure 8: Linear Regression of Alcohol Poisoning onto Age



Figure 9: Deaths Due to Motor Vehicle Accidents and Alcohol Poisoning



Tables

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | No Birthday | | | | Birthday | | |
|  | Linear | Quadratic | Cubic | | Linear | Quadratic | Cubic |
| from21 | 0.00024  (0.000013) | -0.000057  (0.00005) | | 0.000018  (0.00012) | 0.00023  (0.000013) | -0.000065  (0.00005) | -0.000003  (0.00012) |
| over21 | 0.079  (0.013) | 0.12  (0.013) | | 0.11  (0.017) | 0.080  (0.0091) | 0.12  (0.013) | 0.11  (0.017) |
| from21\*over21 | -0.00024  (0.000013) | 0.00007  (0.00005) | | 0.000014  (0.00012) | -0.00024  (0.000013) | 0.000078  (0.000051) | 0.000034  (0.00012) |
| from21\*from21 |  | -0.0\*  (0.0)\* | | -0.0\*  (0.0)\* |  | -0.0\*  (0.0)\* | -0.0\*  (0.0)\* |
| from21\*from21\*over21 |  | 0.0\*  (0.0)\* | | 0.0\*  (0.0)\* |  | 0.0\*  (0.0)\* | 0.0\*  (0.0)\* |
| from21\*from21\*from21 |  |  | | 0.0\*  (0.0)\* |  |  | 0.0\*  (0.0)\* |
| from21\*from21\*from21\*over21 |  |  | | -0.0\*  (0.0)\* |  |  | -0.0\*  (0.0)\* |
| Intercept | 0.59  (0.0079) | 0.55  (0.011) | | 0.55  (0.014) | 0.59  (0.0079) | 0.54  (0.011) | 0.55  (0.014) |
| Birthday |  |  | |  | 0.055  (0.082) | 0.11  (0.083) | 0.10  (0.083) |

Table 1: Regression Table with Estimates of Changes in Drinking

\*0.0 is given when the value has more than six zeros before a significant figure

Table 2: Regression Table with Estimates of Changes in All Deaths

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | No Birthday | | | Birthday | | |
|  | Linear | Quadratic | Cubic | Linear | Quadratic | Cubic |
| from21 | 0.083  (0.82) | -0.83  (3.29) | -5.63  (8.40) | 0.083  (0.79) | -0.83  (3.23) | -5.63  (8.31) |
| over21 | 7.66  (1.32) | 9.55  (1.99) | 11.26  (2.70) | 6.88  (1.33) | 8.06  (2.17) | 8.72  (3.26) |
| from21\*over21 | -3.60  (1.16) | -6.02  (4.65) | -6.79  (11.89) | -3.01  (1.16) | -3.08  (4.94) | 2.38  (13.53) |
| from21\*from21 |  | -0.84  (1.62) | -6.92  (9.90) |  | -0.84  (1.59) | -6.92  (9.79) |
| from21\*from21\*over21 |  | 2.90  (2.28) | 16.05  (13.99) |  | 1.68  (2.37) | 7.00  (15.35) |
| from21\*from21\*from21 |  |  | -2.06  (3.30) |  |  | -2.06  (3.26) |
| from21\*from21\*from21\*over21 |  |  | -0.33  (4.66) |  |  | 2.31  (5.00) |
| Intercept | 93.62  (0.93) | 93.07  (1.40) | 92.28  (1.91) | 93.62  (0.90) | 93.07  (1.38) | 92.28  (1.89) |
| Birthday |  |  |  | 4.86  (2.40) | 4.30  (2.72) | 4.40  (3.23) |

Table 3: Regression Table with Estimates of Changes in Motor Vehicle Accidents

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | No Birthday | | | Birthday | | |
|  | Linear | Quadratic | Cubic | Linear | Quadratic | Cubic |
| from21 | -2.57  (0.47) | -2.93  (1.91) | -5.50  (4.88) | -2.57  (0.46) | -2.93  (1.86) | -5.50  (4.78) |
| over21 | 4.35  (0.75) | 4.66  (1.15) | 5.68  (1.57) | 4.15  (0.77) | 3.65  (1.25) | 3.91  (1.87) |
| from21\*over21 | -1.16  (0.66) | -0.82  (2.71) | -1.83  (6.90) | -0.87  (0.67) | 1.18  (2.84) | 4.56  (7.79) |
| from21\*from21 |  | -0.19  (0.94) | -3.43  (5.75) |  | -0.19  (0.91) | -3.44  (5.63) |
| from21\*from21\*over21 |  | 0.20  (1.33) | 7.97  (8.13) |  | -0.64  (1.37) | 1.67  (8.83) |
| from21\*from21\*from21 |  |  | -1.10  (1.92) |  |  | -1.10  (1.88) |
| from21\*from21\*from21\*over21 |  |  | -0.43  (2.71) |  |  | 1.41  (2.88) |
| Intercept | 29.93  (0.53) | 29.81  (0.82) | 29.39  (1.11) | 29.93  (0.52) | 29.81  (0.79) | 29.39  (1.09) |
| Birthday |  |  |  | 2.38  (1.38) | 2.94  (1.56) | 3.07  (1.86) |

Table 4: Regression Table with Estimates of Changes in Alcohol Poisoning

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | No Birthday | | | Birthday | | |
|  | Linear | Quadratic | Cubic | Linear | Quadratic | Cubic |
| from21 | 0.20 (0.093) | -0.33  (0.37) | 0.68  (0.84) | 0.20 (0.079) | -0.33  (0.31) | 0.68  (0.77) |
| over21 | 0.44  (0.15) | 0.75  (0.22) | 0.97  (0.27) | 0.28  (0.13) | 0.37  (0.21) | 0.47  (0.30) |
| from21\*over21 | -0.35  (0.13) | -0.21  (0.52) | -3.58  (1.18) | -0.23  (0.12) | 0.53  (0.48) | -1.78  (1.25) |
| from21\*from21 |  | -0.27  (0.18) | 1.01  (0.99) |  | -0.27  (0.15) | 1.01  (0.90) |
| from21\*from21\*over21 |  | 0.46  (0.25) | 2.19  (1.39) |  | 0.16  (0.23) | 0.41  (1.42) |
| from21\*from21\*from21 |  |  | 0.43  (0.33) |  |  | 0.43  (0.30) |
| from21\*from21\*from21\*over21 |  |  | -1.44  (0.46) |  |  | -0.92  (0.46) |
| Intercept | 1.43  (0.11) | 1.25  (0.16) | 1.42  (0.19) | 1.43  (0.09) | 1.25  (0.13) | 1.42  (0.17) |
| Birthday |  |  |  | 1.00  (0.24) | 1.08  (0.26) | 0.87  (0.30) |

Works Cited

Angrist, Joshua, et al. “Identification of Causal Effects Using Instrumental Variables.” *Journal of the American Statistical Association*, Vol. 91, No. 434. June 1996. Print. 16 Mar. 2013.

“About the Heath Interview Survey.” *Centers of Disease Control and Prevention*, 13 June 2012. Web. 7 Mar. 2013. <http://www.cdc.gov/nchs/nhis/about\_nhis.htm>.

“National Vital Statistics System.” *Centers of Disease Control and Prevention*, 4 Mar. 2014. Web. 17 Mar. 2013. <http://www.cdc.gov/nchs/nvss.htm>.