

# Regression Discontinuity Design: Minimum Legal Drinking Age

Luca Bella

2026-01-16

## Table of contents

<b>Introduction</b>	<b>1</b>
<b>Setup</b>	<b>2</b>
Reproducibility . . . . .	2
Packages and data download . . . . .	2
<b>Data import</b>	<b>2</b>
<b>Methods</b>	<b>3</b>
Regression Discontinuity Design . . . . .	3
Estimation and Inference . . . . .	3
<b>Results</b>	<b>4</b>
<b>Robustness</b>	<b>5</b>
<b>Conclusion</b>	<b>8</b>
<b>References</b>	<b>9</b>

## Introduction

In this project we estimate the causal effect of reaching the Minimum Legal Drinking Age (MLDA = 21) on mortality using a sharp Regression Discontinuity Design (RDD). The institutional rule that legal access to alcohol begins exactly at age 21 generates a discontinuity that can be exploited for causal inference.

MAIN ASSUMPTION: Identification of the causal effect relies on the assumption that potential outcomes are smooth functions of age in the absence of treatment. Under this assumption, any discontinuity in mortality outcomes at age 21 can be attributed to legal access to alcohol rather than to other age-related factors.

## Setup

### Reproducibility

**IMPORTANT:** This project uses the `renv` package to ensure full reproducibility. To reproduce the analysis, run `renv::restore()` in the Console after cloning the repository.

### Packages and data download

Load packages, download and store data in a new directory.

```
suppressPackageStartupMessages({  
  library(tidyverse)  
  library(haven)  
  library(rdrobust)  
  library(rddensity)  
})  
  
dir.create("data/raw", recursive = TRUE, showWarnings = FALSE)  
  
url <- "https://pup-assets.s3.amazonaws.com/public/resources/9780691152844/Chapter-4-Table-4-  
zip <- "data/raw/mlda.zip"  
  
if (!file.exists(zip)) download.file(url, zip, mode = "wb")  
unzip(zip, exdir = "data/raw")
```

## Data import

We use age-in-month mortality data from Carpenter and Dobkin (2011), distributed with *Mostly Harmless Econometrics*.

We'll apply two main modifications here. The dataset includes both raw mortality outcomes and smoothed series (denoted by variables ending in “fitted”) that were constructed for visualization purposes in *Mostly Harmless Econometrics*. All analyses in this paper use the raw

mortality measures; the smoothed series are excluded from estimation. The running variable is age in years and we center it at the legal drinking age cutoff for interpretability purposes.

```
mlda <- read_dta("data/raw/AEJfigs.dta") # mlda stands for "minimum legal drinking age (data)

# we import the dataset which is in .dta format which is a Stata data file format.
mlda <- mlda %>%
  mutate(
    running = agecell - 21
  ) %>%
  select(-ends_with("fitted")) %>%
  filter(between(running, -2, 2))
```

## Methods

### Regression Discontinuity Design

We estimate the causal effect of legal access to alcohol using a sharp regression discontinuity design (RDD), exploiting the deterministic change in legal drinking status at age 21. Age, measured in months and centered at the cutoff, serves as the running variable. Treatment assignment changes discontinuously at the cutoff, creating a quasi-experimental comparison between individuals just below and just above age 21.

Identification relies on two key assumptions. First, potential outcomes must evolve smoothly in age in the absence of treatment, so that any discontinuity in mortality at the cutoff can be attributed to legal drinking access. Second, individuals must not be able to precisely manipulate the running variable. Age is fixed and not subject to individual control, making this assumption highly credible in the MLDA setting.

The estimand is the local average treatment effect at the cutoff, defined as the difference in expected outcomes immediately above and below age 21. As with all RDD estimates, this effect is local and should not be extrapolated far from the cutoff.

### Estimation and Inference

We implement local linear regressions on either side of the cutoff, which are standard in RDD applications due to their favorable boundary properties. Estimation focuses on observations within a data-driven bandwidth around the cutoff, balancing bias and variance.

Inference follows the robust bias-corrected procedure developed by Calonico, Cattaneo, and Titiunik (2014). Local polynomial estimators near the cutoff suffer from boundary bias because observations are only available on one side of the threshold. Robust bias-corrected inference

explicitly estimates this bias, adjusts the point estimate accordingly, and inflates standard errors to account for the additional uncertainty introduced by bias estimation. This approach delivers confidence intervals with correct coverage and is now standard practice in applied regression discontinuity analyses.

All estimates are computed using the `rdrobust` package, which implements mean squared error-optimal bandwidth selection and robust bias-corrected confidence intervals. Covariates are not required for identification in a sharp RDD and are omitted to preserve transparency and interpretability.

## Results

```
rd_all <- rdrobust(
  y = mlda$all,
  x = mlda$running,
  c = 0
)

summary(rd_all)
```

Sharp RD estimates using local polynomial regression.

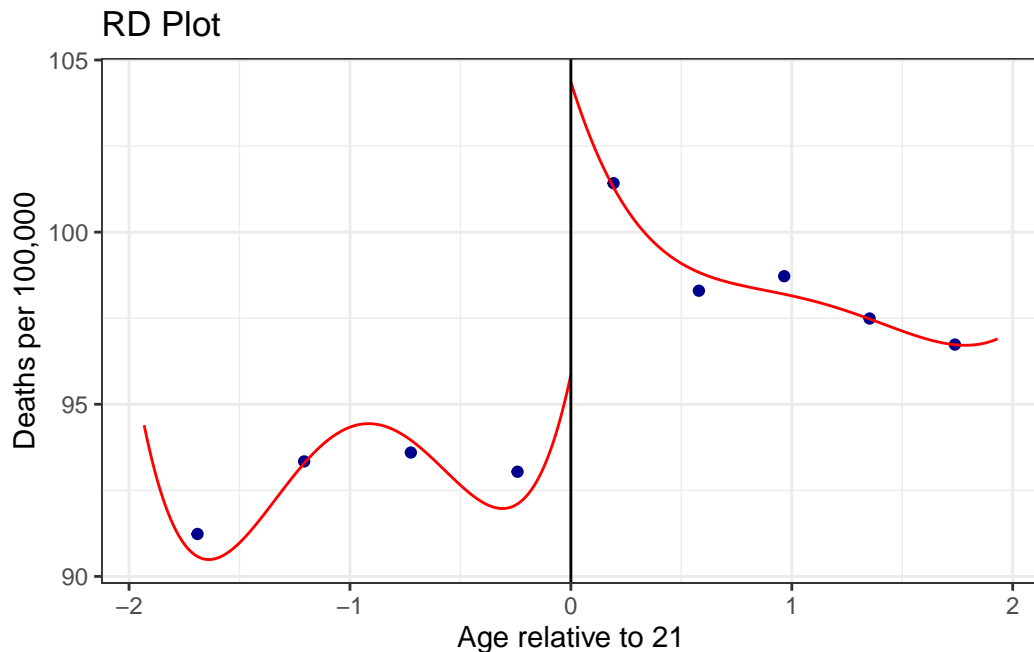
Number of Obs.	48	
BW type	mserd	
Kernel	Triangular	
VCE method	NN	
Number of Obs.	24	24
Eff. Number of Obs.	6	6
Order est. (p)	1	1
Order bias (q)	2	2
BW est. (h)	0.493	0.493
BW bias (b)	0.780	0.780
rho (h/b)	0.632	0.632
Unique Obs.	24	24

```
=====
              Point      Robust Inference
              Estimate      z      P>|z|      [ 95% C.I. ]
-----
```

RD Effect	9.595	2.205	0.027	[1.077 , 18.300]
=====				

The regression discontinuity estimates indicate a statistically significant increase in all-cause mortality at age 21. The estimated discontinuity is economically meaningful, suggesting that legal access to alcohol leads to a discrete rise in death rates among individuals just above the drinking age threshold. Because individuals just below and just above age 21 are otherwise comparable, this discontinuity can be interpreted as a causal effect of legal alcohol access on mortality in a local neighborhood around the cutoff.

```
rdplot(mlda$all, mlda$running, c = 0,
       x.label = "Age relative to 21",
       y.label = "Deaths per 100,000")
```



## Robustness

```
# Density test
dens <- rddensity(mlda$running, c = 0)
summary(dens)
```

Manipulation testing using local polynomial density estimation.

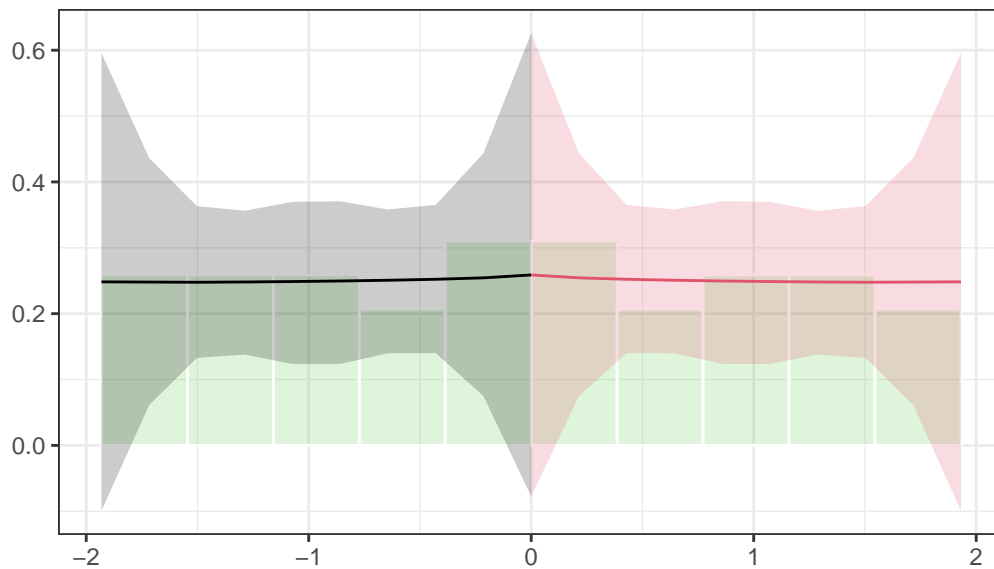
Number of obs =	50	
Model =	unrestricted	
Kernel =	triangular	
BW method =	estimated	
VCE method =	jackknife	
c = 0	Left of c	Right of c
Number of obs	25	25
Eff. Number of obs	23	23
Order est. (p)	2	2
Order bias (q)	3	3
BW est. (h)	1.767	1.767
Method	T	P >  T
Robust	0	1

P-values of binomial tests ( $H_0: p=0.5$ ).

Window Length / 2	<c	>=c	P> T
1.521	20	20	1.0000
1.548	20	20	1.0000
1.575	20	20	1.0000
1.603	21	21	1.0000
1.630	21	21	1.0000
1.658	21	21	1.0000
1.685	22	22	1.0000
1.712	22	22	1.0000
1.740	22	22	1.0000
1.767	23	23	1.0000

```
# the following commands are to avoid excess space below graphs when rendered
#| fig-cap: false
#| fig-width: 4
#| fig-asp: 0.25
#| fig-pos: "H"

# Density plot
rdplotdensity(mlda$running, rdd = dens)
```



\$Est1

Call: lpdensity

Sample size	25
Polynomial order for point estimation (p=)	2
Order of derivative estimated (v=)	1
Polynomial order for confidence interval (q=)	3
Kernel function	triangular
Scaling factor	0.510204081632653
Bandwidth method	user provided

Use summary(...) to show estimates.

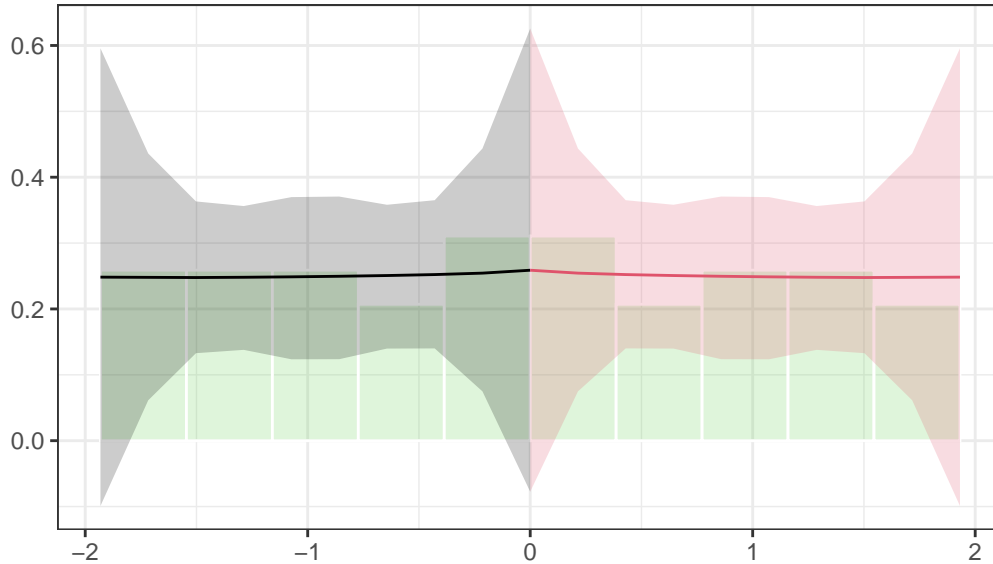
\$Estr

Call: lpdensity

Sample size	25
Polynomial order for point estimation (p=)	2
Order of derivative estimated (v=)	1
Polynomial order for confidence interval (q=)	3
Kernel function	triangular
Scaling factor	0.510204081632653
Bandwidth method	user provided

Use summary(...) to show estimates.

\$Estplot



We formally test for manipulation of the running variable at the minimum legal drinking age cutoff using local polynomial density estimation. The results show no evidence of a discontinuity in the density of age at the cutoff. The robust test statistic is zero with a p-value of one, and binomial tests across a range of bandwidths yield identical counts on either side of the threshold. These findings indicate that individuals are unable to precisely sort around the legal drinking age, supporting the validity of the regression discontinuity design.

The graph shows the robust local polynomial density estimates of the running variable around the legal drinking age cutoff. The estimated densities evolve smoothly on both sides of the threshold, with no visible discontinuity at age 21. Confidence bands overlap at the cutoff, providing graphical support for the formal manipulation test. Together, these results indicate no evidence of sorting or manipulation of the running variable, supporting the validity of the regression discontinuity de

## Conclusion

Mortality increases discretely at age 21, driven largely by external causes. The results are consistent with increased risky behavior following legal access to alcohol.



## References

Angrist, J. D., & Pischke, J.-S. (2009). *Mostly Harmless Econometrics*. Princeton University Press. Carpenter, C., & Dobkin, C. (2011). The Minimum Legal Drinking Age and Public Health. *Journal of Economic Perspectives*. Calonico, Calonico, S., Cattaneo, M. D., & Titiunik, R. (2014). Robust nonparametric confidence intervals for regression-discontinuity designs. *Econometrica*, 82(6), 2295–2326.