

回帰不連続デザイン

講師：遠山祐太

最終更新：2025-01-20

はじめに

はじめに

- **回帰不連続デザイン (regression discontinuity design; RDD)**
 - 処置状況の不連続変化を使って因果効果を推定する。
- 例
 - 大学入試における試験成績の閾値 (threshold)
 - 年齢による政策の適用可否
 - 地区の地理的な境界

例：時短対象

テレ朝news

FOLLOW US

テレ朝newsを検索

トップ | 新型コロナ | 速報 | 社会 | 政治 | 経済 | 国際 | スポーツ

三鷹駅の「南北問題」 “明暗”分かれ飲食店は…

[2021/04/13 19:45]

三鷹駅「南北問題」で明暗 閑散、時短対象の武蔵野市

武蔵野市

北口・対象内

JR三鷹駅

南口・対象外

三鷹市

本日の流れと参考文献

- 流れ
 - フレームワーク
 - ケース例：飲酒可能年齢と死亡率
 - RDDにおける仮定の吟味
- 付録資料 (アップデートしていない点に留意)
 - 付録1 : Shigeoka (2014,AER)
 - 付録2 : Rでの実践 (英語のみ)
- 参考文献
 - Angrist and Pischke "*Mostly harmless econometrics*" Chapter 6
 - R のパッケージ:[rdrobust](#)
 - パッケージの開発者たちによるサーベイ文献 (英語) もおすすめ。

フレームワーク

フレームワーク

- Y_i : 個人 i の観測できるアウトカム
- **潜在アウトカム**
 - Y_{1i} : i が処置を施されるときのアウトカム
 - Y_{0i} : i が処置を施されないときのアウトカム
- **シャープな (sharp) 回帰不連続デザイン**では、処置の状況 D_i は一意に定まる。

$$D_i = \mathbf{1}\{W_i \geq \bar{W}\}$$

- W_i : 割当変数 (running variable)
- \bar{W} : 閾値 (threshold)
- 確率的な割り当ても同様に考えられ、**ファジーな (fuzzy) 回帰不連続デザイン**という。

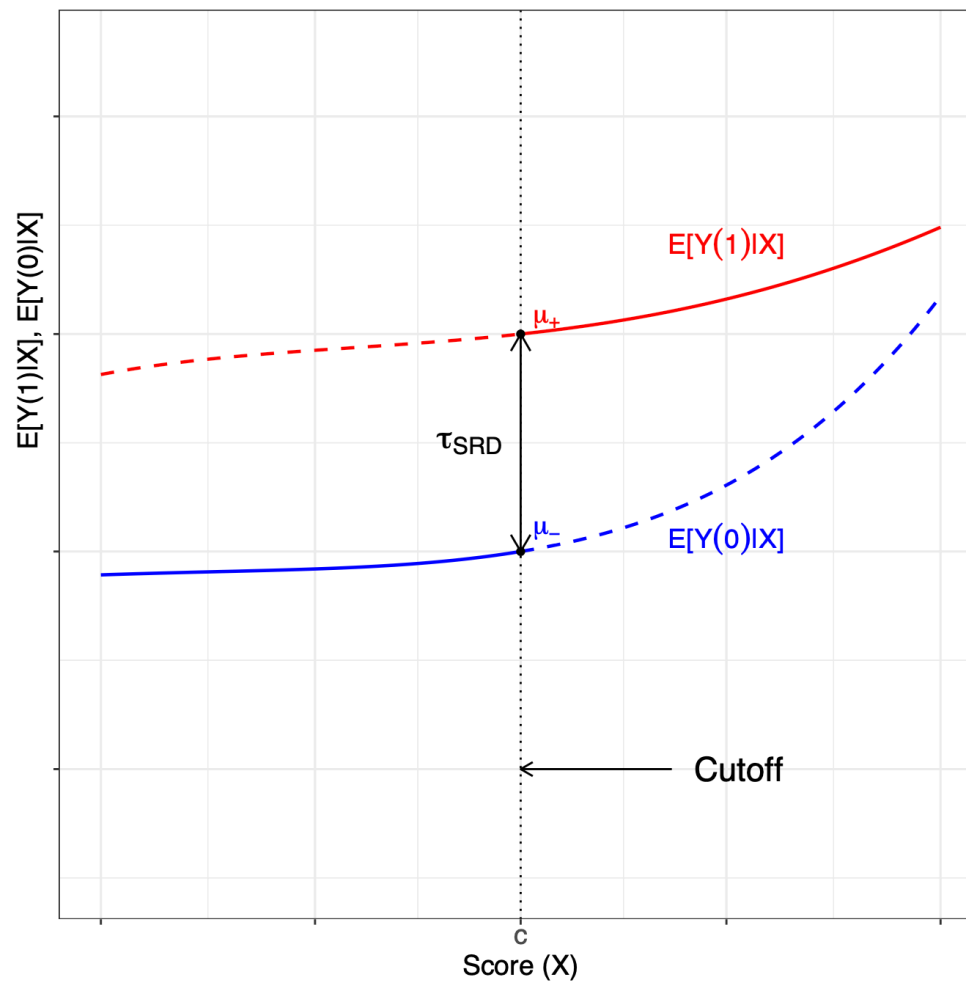
例 : Incumbency Advantage

- 候補者が二人の選挙を考える。
 - D_i : 現職ダミー
 - Y_i : 当選ダミー
 - W_i : 前回の選挙における得票率
- 現職であれば

$$D_i = \mathbf{1}\{W_i \geq 0.5\}$$

- 回帰不連続デザインのアイデア：
 - 得票率 51% で当選する場合を考える。
 - この当選者は得票率 49% で落選した候補者と似ている。(RDの主要な仮定)
 - これらの候補者に注目すれば、 D_i はあたかも無作為に割り当てられているといえる。

RDDの考え方 (Score X = Running variable W)



続・フレームワーク

- $D_i = \mathbf{1}\{W_i \geq \bar{W}\}$ であれば、条件付き無視可能性

$$(Y_{1i}, Y_{0i}) \perp D_i | W_i$$

が成り立つ。

- しかし、共通サポートの仮定は満たされない。なぜならば

$$P(D_i = 1 | W_i = w) = \begin{cases} 1 & \text{if } w \geq \bar{W} \\ 0 & \text{if } w < \bar{W} \end{cases}$$

- いかなる w についても、処置群・対照群を同時に観察することはできない。
- 処置群と統制群を比較するには、閾値周辺における外挿に頼らなければならない。

線形手法による外装

- 次の定式化を考えよう。

$$Y_{1i} = \rho + Y_{0i}$$
$$E[Y_{0i}|W_i = w] = \alpha_0 + \beta_0 w$$

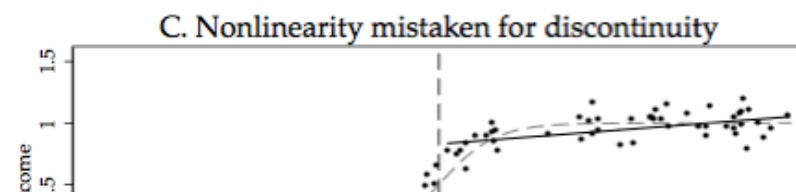
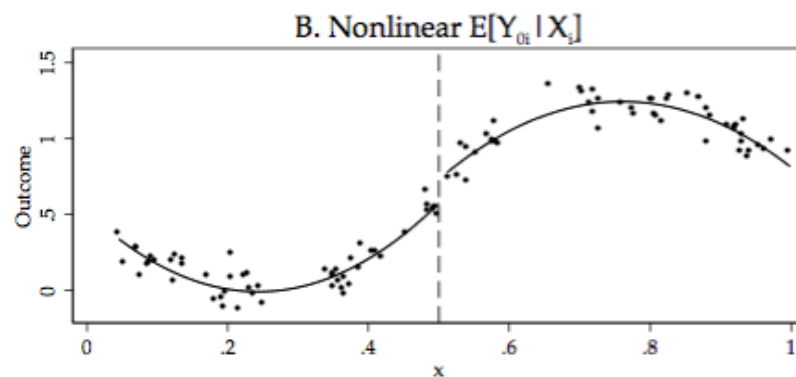
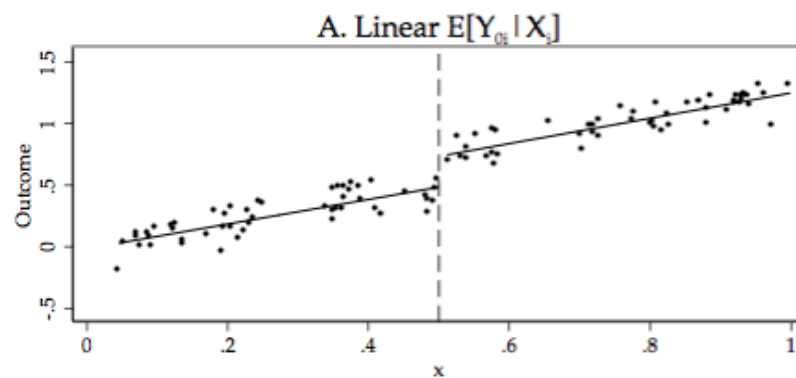
- これは回帰

$$Y_i = \alpha + \beta W_i + \rho D_i + \eta_i$$

を導く。ここで ρ は因果効果である。

- もし $E[Y_{0i}|W_i = w]$ が非線形であればどうなる？

線形・非線形の比較



より一般的な手法

- W_i に非線形な効果を許す。

$$Y_i = f(W_i) + \rho \mathbf{1}\{W_i \geq \bar{W}\} + \eta_i$$

- 関数 $f(\cdot)$ は p 次多項式とする。

$$f(W_i) = \beta_1 W_i + \beta_2 W_i^2 + \cdots + \beta_p W_i^p$$

回帰での実装

- 式

$$E[Y_{0i}|W_i = w] = f_0(W_i - \bar{W})$$

$$E[Y_{1i}|W_i = w] = \rho + f_1(W_i - \bar{W})$$

- 回帰式は

$$Y_i = \alpha + \beta_{01}\tilde{W}_i + \cdots + \beta_{0p}\tilde{W}_i^p \\ + \rho D_i + \beta_1^* D_i \tilde{W}_i + \cdots + \beta_p^* D_i \tilde{W}_i^p + \eta_i$$

- $\tilde{W}_i = W_i - \bar{W}$ は正規化である。
- 回帰の際は、閾値まわりの標本に注目しなければならない。

推定に関する実践

- 実行に際しては選ぶべき要素が多い。
 - どの程度閾値に近いサンプルを選ぶか（バンド幅）
 - 多項式の次数、
 - ローカル線形回帰のカーネルの種類
 - その他もろもろ
- Rのパッケージの `rdrobust` がオススメ。マニュアル（英語）も提供されている。
 - 分析で実際にRDを行う人は必読

フォーマルな識別の議論

- 連続性の仮定： $E[Y_{1i}|W_i = w]$ と $E[Y_{0i}|W_i = w]$ は 閾値 $w = \bar{W}$ で連続である。
- この仮定のもとで、閾値におけるATT $E[Y_{1i} - Y_{0i}|W_i = \bar{W}]$ が識別される。

識別の証明（スケッチ）

- まず以下の展開を考える。

$$E[Y_i|W_i = w] = E[Y_{0i}|W_i = w] + \mathbf{1}\{w \geq \bar{W}\} (E[Y_{1i}|W_i = w] - E[Y_{0i}|W_i = w])$$

- $w \rightarrow \bar{W}$ の上極限をとると

$$\lim_{w \downarrow \bar{W}} E[Y_i|W_i = w] = \lim_{w \downarrow \bar{W}} E[Y_{1i}|W_i = w] = E[Y_{1i}|W_i = \bar{W}]$$

続証明：下極限

- 今度は下極限を考える

$$\lim_{w \uparrow \bar{W}} E[Y_i | W_i = w] = \lim_{w \uparrow \bar{W}} E[Y_{0i} | W_i = w] = E[Y_{0i} | W_i = \bar{W}]$$

二つの極限をまとめると・・・

- まとめます

$$\lim_{w \uparrow \bar{W}} E[Y_i | W_i = w] = \lim_{w \uparrow \bar{W}} E[Y_{0i} | W_i = w] = E[Y_{0i} | W_i = \bar{W}]$$

$$\lim_{w \downarrow \bar{W}} E[Y_i | W_i = w] = \lim_{w \downarrow \bar{W}} E[Y_{1i} | W_i = w] = E[Y_{1i} | W_i = \bar{W}]$$

- よって、

$$E[Y_{1i} - Y_{0i} | W_i = \bar{W}] = \lim_{w \downarrow \bar{W}} E[Y_i | W_i = w] - \lim_{w \uparrow \bar{W}} E[Y_i | W_i = w]$$

- 左辺：閾値での平均処置効果
- 右辺：データから観測できる差分(閾値まわりでの条件付き期待値)

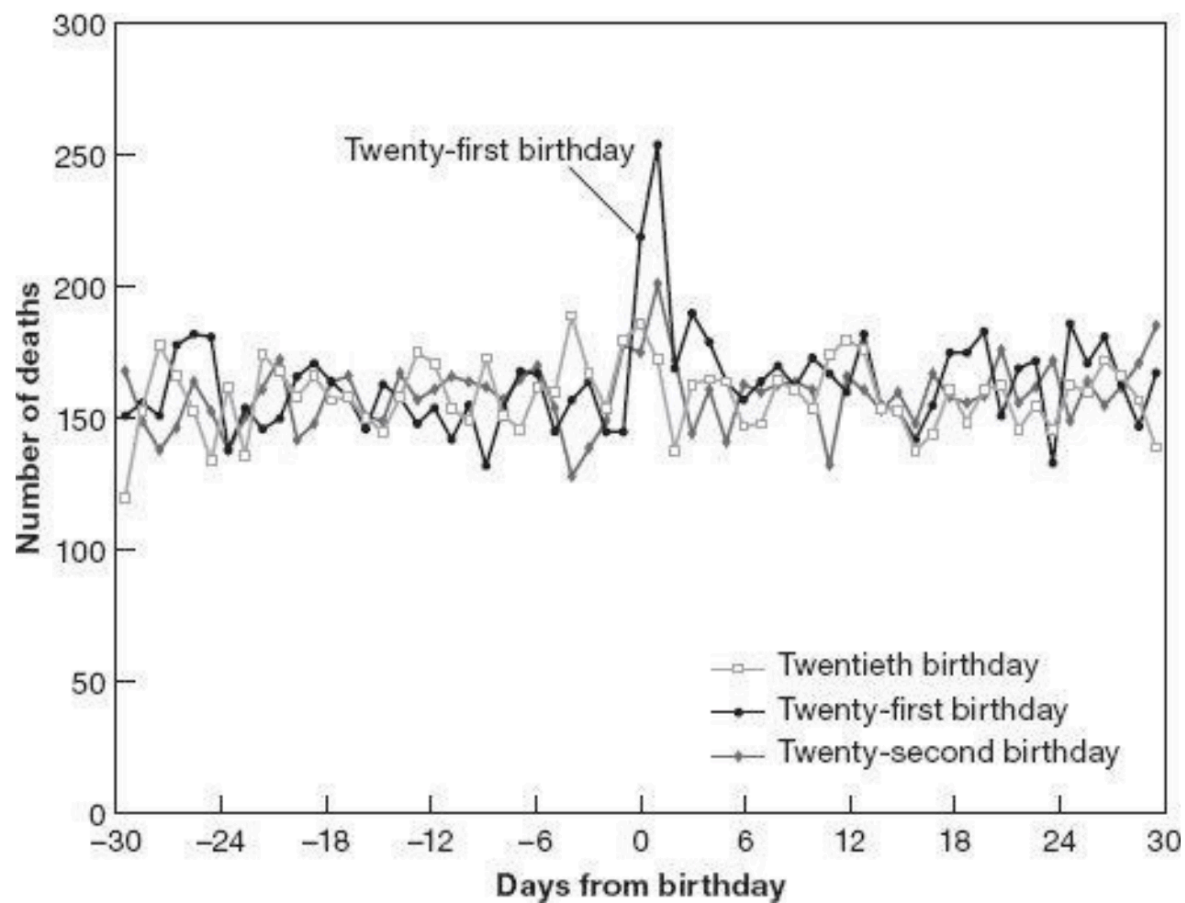
ケース:法定飲酒可能年齢の効果

ケース：法定飲酒可能年齢の効果

- 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.
- 問：アルコール消費が死亡率にどのような影響を与えるか？
- 米国における法定飲酒可能年齢（21歳）による飲酒の有無をRDで分析。
- 以下のスライドにおける図表はAngrist and Pischke "Matering Metrics"より

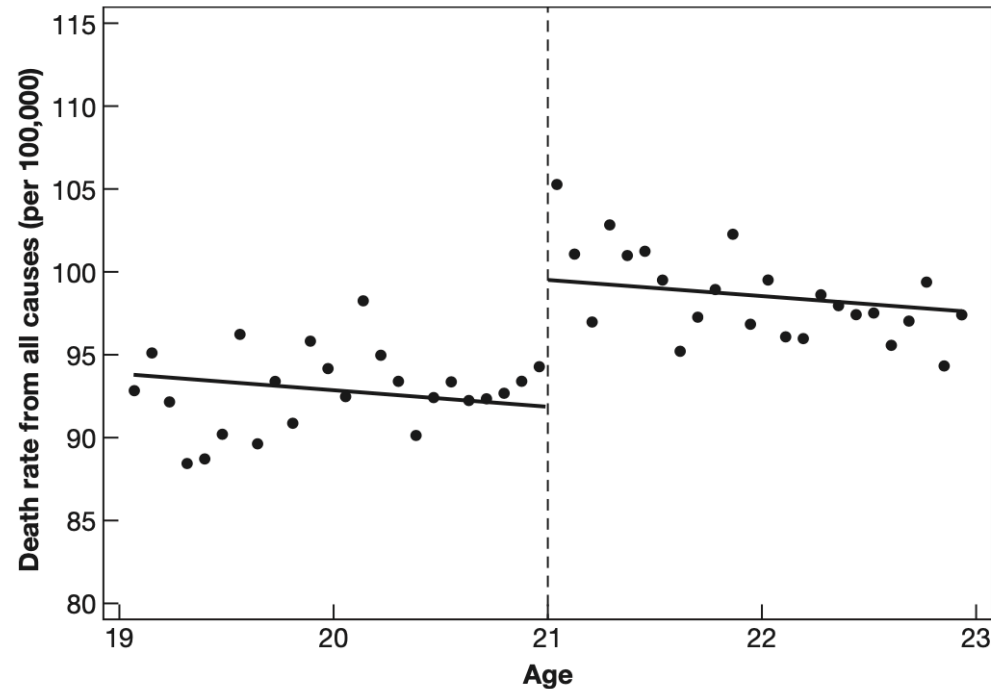
誕生日と死亡件数

FIGURE 4.1
Birthdays and funerals



飲酒可能性が死亡率に与える影響（両側線形）

FIGURE 4.2
A sharp RD estimate of MLDA mortality effects

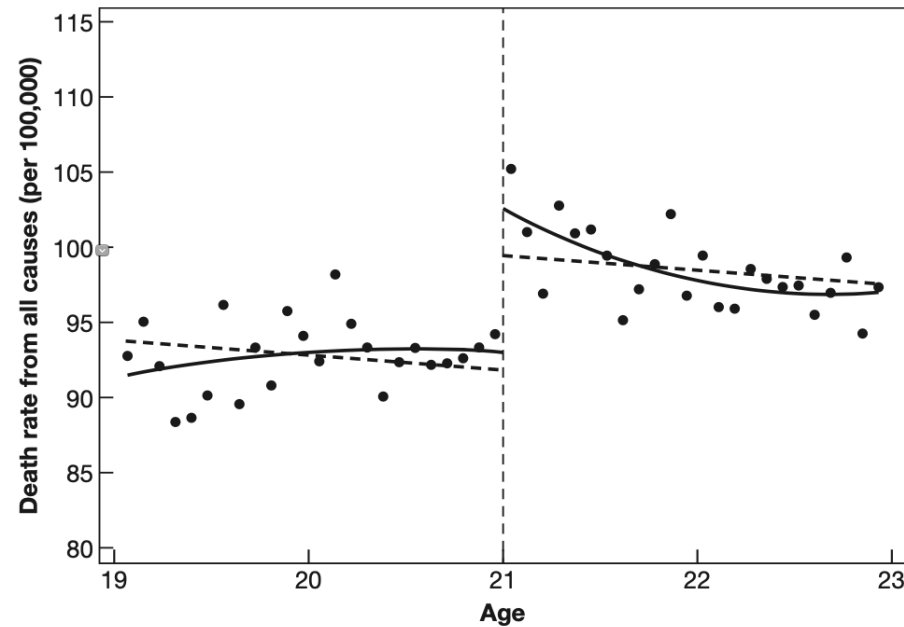


Notes: This figure plots death rates from all causes against age in months. The lines in the figure show fitted values from a regression of death rates on an over-21 dummy and age in months (the vertical dashed line indicates the minimum legal drinking age (MLDA) cutoff).

両側を二次関数でフィットした場合

tel:115%20110%20105%20100

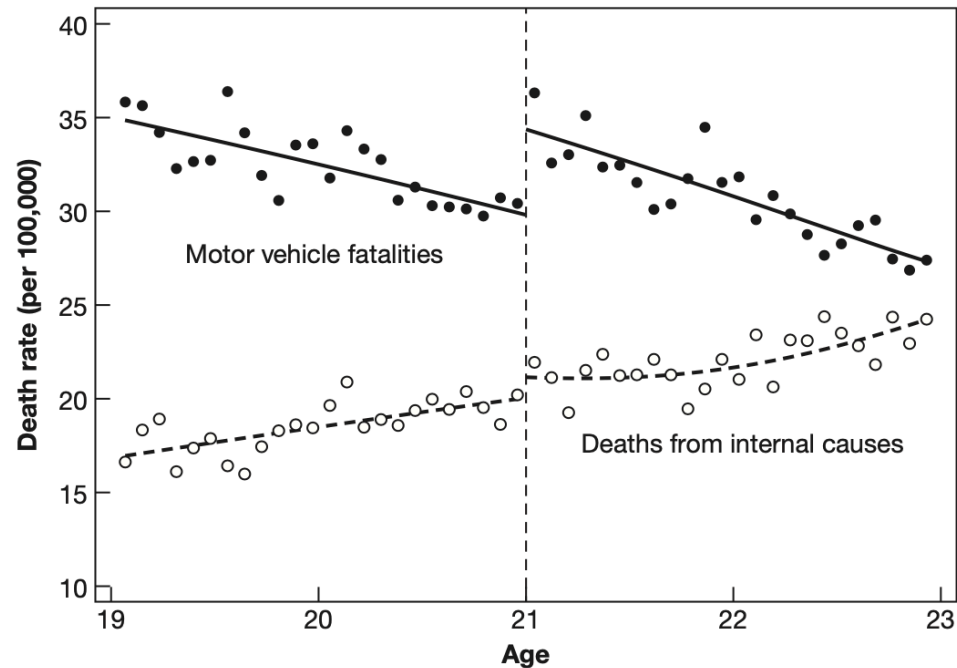
FIGURE 4.4
Quadratic control in an RD design



Notes: This figure plots death rates from all causes against age in months. Dashed lines in the figure show fitted values from a regression of death rates on an over-21 dummy and age in months. The solid lines plot fitted values from a regression of mortality on an over-21 dummy and a quadratic in age, interacted with the over-21 dummy (the vertical dashed line indicates the minimum legal drinking age [MLDA] cutoff).

死亡原因別の効果

FIGURE 4.5
RD estimates of MLDA effects on mortality by cause of death



Notes: This figure plots death rates from motor vehicle accidents and internal causes against age in months. Lines in the figure plot fitted values from regressions of mortality by cause on an over-21 dummy and a quadratic function of age in months, interacted with the dummy (the vertical dashed line indicates the minimum legal drinking age [MLDA] cutoff).

推定結果まとめ

TABLE 4—DISCONTINUITY IN LOG DEATHS AT AGE 21

	(1)	(2)	(3)	(4)
<i>Deaths due to all causes</i>				
Over 21	0.096 (0.018)	0.087 (0.017)	0.091 (0.023)	0.074 (0.016)
Observations	1,460	1,460	1,460	1,458
R^2	0.04	0.05	0.05	
Prob > Chi-Squared		0.000	0.735	
<i>Deaths due to external causes</i>				
Over 21	0.110 (0.022)	0.100 (0.021)	0.096 (0.028)	0.082 (0.021)
Observations	1,460	1,460	1,460	1,458
R^2	0.06	0.08	0.08	
Prob > Chi-Squared		0.000	0.788	
<i>Deaths due to internal causes</i>				
Over 21	0.063 (0.040)	0.054 (0.040)	0.094 (0.053)	0.066 (0.031)
Observations	1,460	1,460	1,460	1,458
R^2	0.10	0.10	0.10	
Prob > Chi-Squared		0.000	0.525	
Covariates	N	Y	Y	N
Quadratic terms	Y	Y	Y	N
Cubic terms	N	N	Y	N
LLR	N	N	N	Y

Notes: See Notes from Table 1. The dependent variable is the log of the number of deaths that occurred x days from the person's twenty-first birthday. External deaths include all deaths with mention of an injury, alcohol use, or drug use. The Internal Death category includes all deaths not coded as external. Please see Web Appendix C for the ICD codes for each of the categories above. The first three columns give the estimates from polynomial regressions on age interacted with a dummy for being over 21.

死亡原因別 (外的要因)

TABLE 5—DISCONTINUITY IN LOG DEATHS AT AGE 21 DUE TO EXTERNAL CAUSES

	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
	Alcohol				Homicide			
Over 21	0.388 (0.119)	0.346 (0.116)	0.406 (0.156)	0.441 (0.117)	0.009 (0.045)	0.002 (0.045)	-0.003 (0.061)	-0.014 (0.041)
Observations	1,460	1,460	1,460	1,458	1,460	1,460	1,460	1,458
R ²	0.03	0.04	0.04			0.01	0.01	0.01
Prob > Chi-Squared		0.000	0.228			0.000	0.495	
	Suicide				Motor vehicle accidents			
Over 21	0.160 (0.059)	0.154 (0.059)	0.135 (0.086)	0.105 (0.045)	0.158 (0.033)	0.143 (0.032)	0.145 (0.044)	0.139 (0.032)
Observations	1,460	1,460	1,460	1,458	1,460	1,460	1,460	1,458
R ²	0.02	0.02	0.02	0.15	0.16	0.16		
Prob > Chi-Squared		0.000	0.892	0.000	0.666			
	Drugs				Other external causes			
Over 21	0.070 (0.081)	0.067 (0.082)	0.004 (0.107)	-0.016 (0.078)	0.087 (0.060)	0.098 (0.059)	0.098 (0.075)	0.074 (0.043)
Observations	1,460	1,460	1,460	1,458	1,460	1,460	1,460	1,458
R ²	0.04	0.04	0.04	0.01	0.01	0.01		
Prob > Chi-Squared		0.000	0.643			0.000	0.877	
Covariates	N	Y	Y	N	N	Y	Y	N
Quadratic terms	Y	Y	Y	N	Y	Y	Y	N
Cubic terms	N	N	Y	N	N	N	Y	N
LLR	N	N	N	Y	N	N	N	Y

Notes: See Notes from Table 4. There are 276 observations where there are N deaths coded as due to alcohol; for this variable 0.5 was added to the dependent variable before taking the log. There are 15 observations where there are N deaths coded as due to drug use; for this variable 0.5 was added to the count before taking the log.

仮定の妥当性

回帰不連続デザインにおける仮定の妥当性

- 主要な仮定 1 : SUTVA (閾値を越えた処置の波及がない)
- 主要な仮定 2 : 閾値での潜在アウトカム連続性

SUTVAの満たされない例

酒飲める店求め川を渡る 宣言下の東京から隣県へ

社会・暮らし

2021年4月26日 22:00 (2021年4月26日 22:28更新) [有料会員限定]

保存済み

あ ら 印刷 送信 共有 ツイット Facebook 共有



JR川口駅周辺の多くの居酒屋は営業中で客の姿があった（25日、埼玉県川口市）

新型コロナウイルス対策の緊急事態宣言が発令された東京都内で飲食店の休業や酒類の提供停止が広がり、宣言対象外のエリアに人が流れている。都は不要不急の移動を控えるよう求めているが、埼玉県や神奈川県でかえって人出が増えた街もある。飲食の場での感染拡大を抑える対策が骨抜きになり、周辺地域に感染を広げるリスクがある。

連続性

- 主要な仮定 2 : $E[Y_{1i}|W_i = w]$ と $E[Y_{0i}|W_i = w]$ がともに閾値 $w = \bar{W}$ で連続である。
- 閾値以前の Y_{1i} を観測できないため、この仮定を直接検定することはできない。
- 仮定をサポートするために用いられる手法
 1. 共変量テスト
 2. 密度テスト (割当変数の集積がない)

共変量テスト

- 回帰不連続デザインの根底にあるアイデア： 閾値のちょうど前後のアウトカムを比較することで、条件付き分布の連続性の仮定によって似通っている処置群と対照群の比較ができる。
- もしこれが妥当な比較なら、閾値の前後で共変量 X も滑らかに変化するはずである。

共変量テスト

- 回帰不連続デザインを共変量 X についておこなう。
- もし不連続なら、 W に条件付けた Y もまた、不連続だと考えられる。
- もし X が Y に直接効果をもつなら、 $E[Y_i|W]$ の \bar{W} での不連続性は処置効果と交絡する。
- 例：
 - Y : 労働時間
 - D : 65 歳以上割引
 - W : 年齢
 - X : 社会保障給付 (不労所得)

割当の操作と集積点(Bunching)

- 制度の詳細が知られている場合に、割当変数に意図的な操作が行われる可能性。
- 例：もし、全国统一テストにおいて50点以下の学校が機能不全のレッテルを貼られるならば、学校は平均点を操作するインセンティブがある。
- 結果、多くの学校の平均点が**50 点のすぐ上に集まる**という不自然な分布になる可能性。

密度テスト

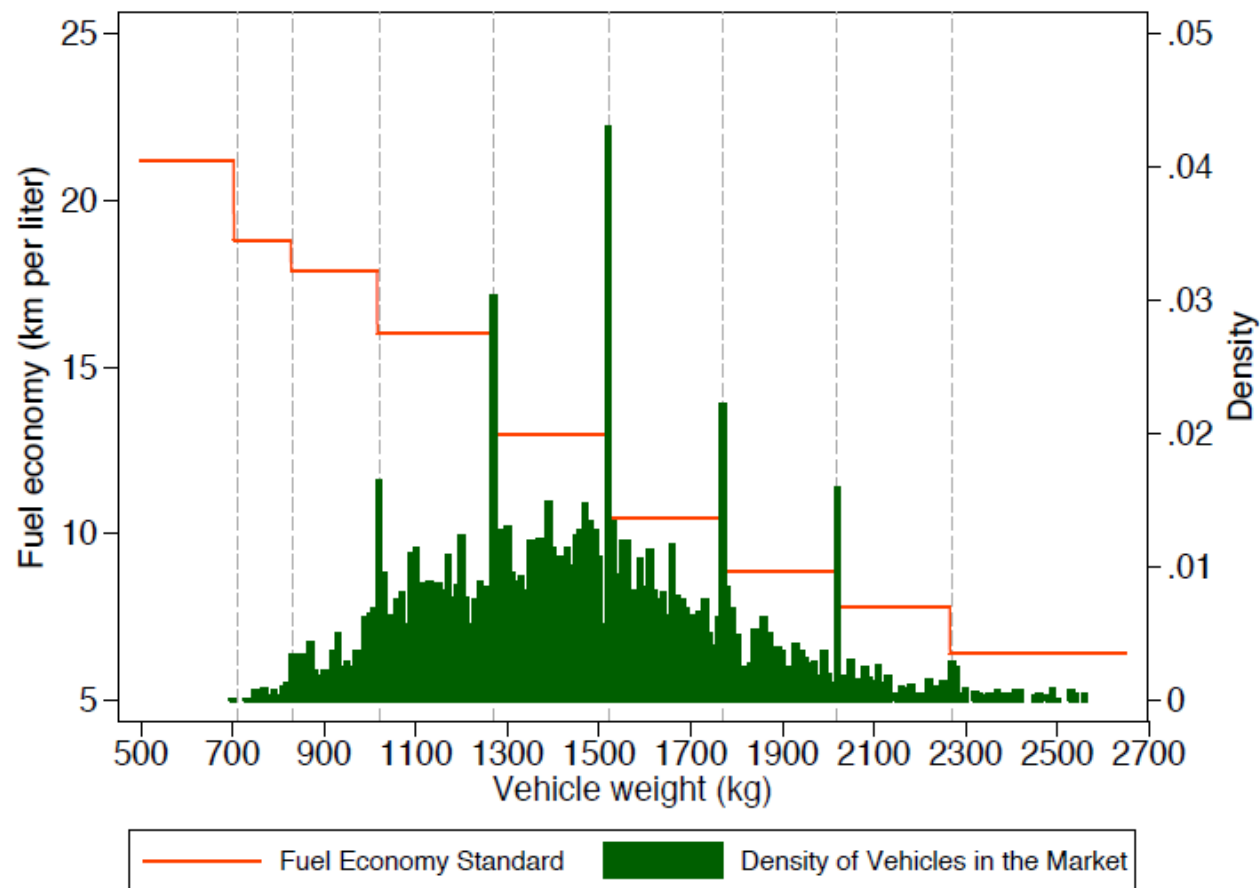
- この場合、閾値まわりでの**集積 (bunching)** が起こっている。
 - エージェントは閾値まわりで処置の割り当てを「操作」している。
 - W_i の密度が \bar{W} で不連続になる。
- 結果、 $E[Y_{1i}|W_i = w]$ も不連続になると考えられる。
- McCrary (2008) の検定： 割当変数 W_i の密度が \bar{W} で連続であるかを検定

おまけ：集積点分析 (Bunching Analysis)

- 集積があるとRDの適用は難しくなるが、それ自体は非常に面白い現象であり、多くの論文が書かれている。
- 例：
 - Ito and Salee (2018): 自動車の燃費規制と重量。詳しくは伊藤「データ分析の力」
 - [Shigeoka \(2016\)](#) School Entry Cutoff: 4月1日生まれと2日生まれで学年が分かれる。1日生まれと2日生まれで断絶が見られる。
 - [近藤・深井\(2023\)](#): 「年収の壁」 (扶養基準である103万・130万)

例 : Ito and Salee (2018)

Panel A. Years 2001 to 2008 (Old Fuel-Economy Standard Schedule)



例 : Shigeoka (2016) 学年の断絶

Figure 2: Mean daily number of births around April 2
A. Raw data

