Regressão Discontínua: sharp

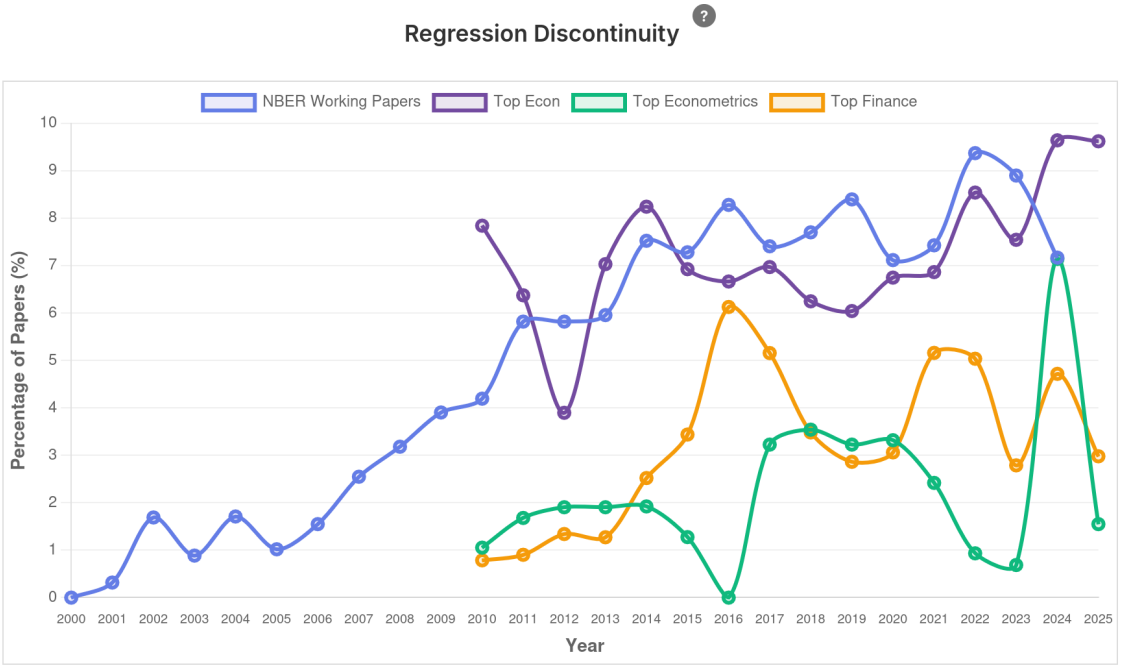
IBM0288 - 2026.2

Prof. Raphael Gouvea

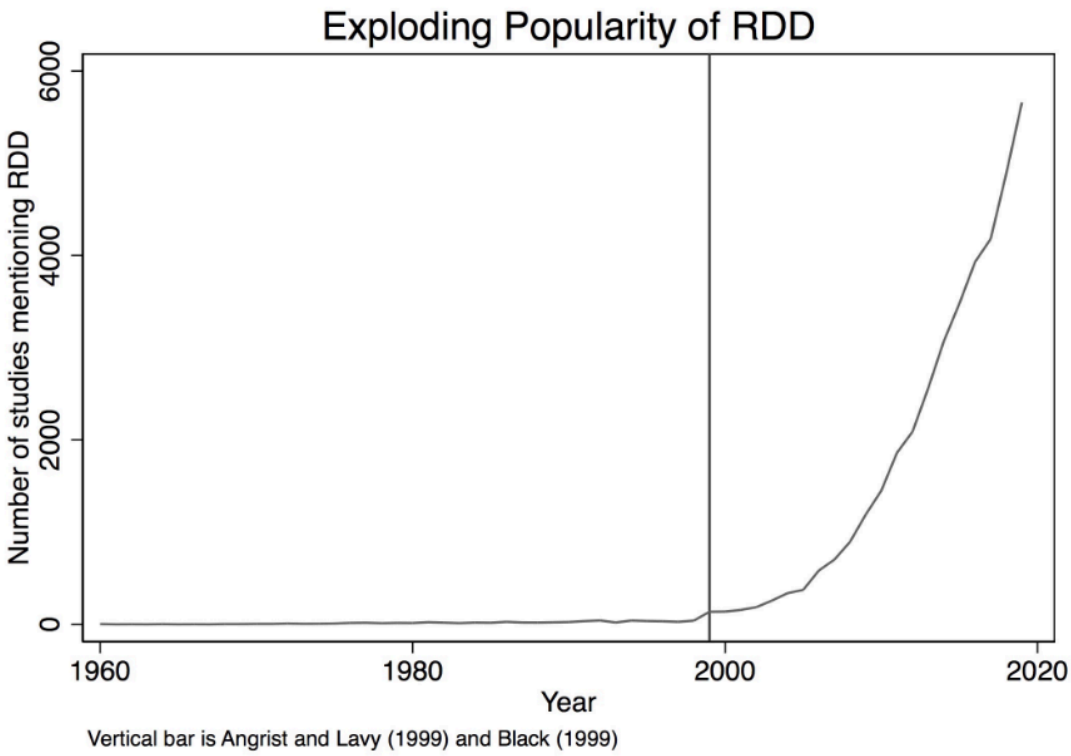
0.1 O que é regressão discontínua?

- RDD (*regression discontinuity design*) é um tipo particular de desenho de pesquisa bastante popular
- Foi inventado por Donald Campbell (1960), um psicólogo da educação. Essa estratégia empírica ficou dormente até cerca de 1999 quando foi "redescoberta" por Angrist, Lavy e Black.

0.2 Motivação I



0.3 Motivação II



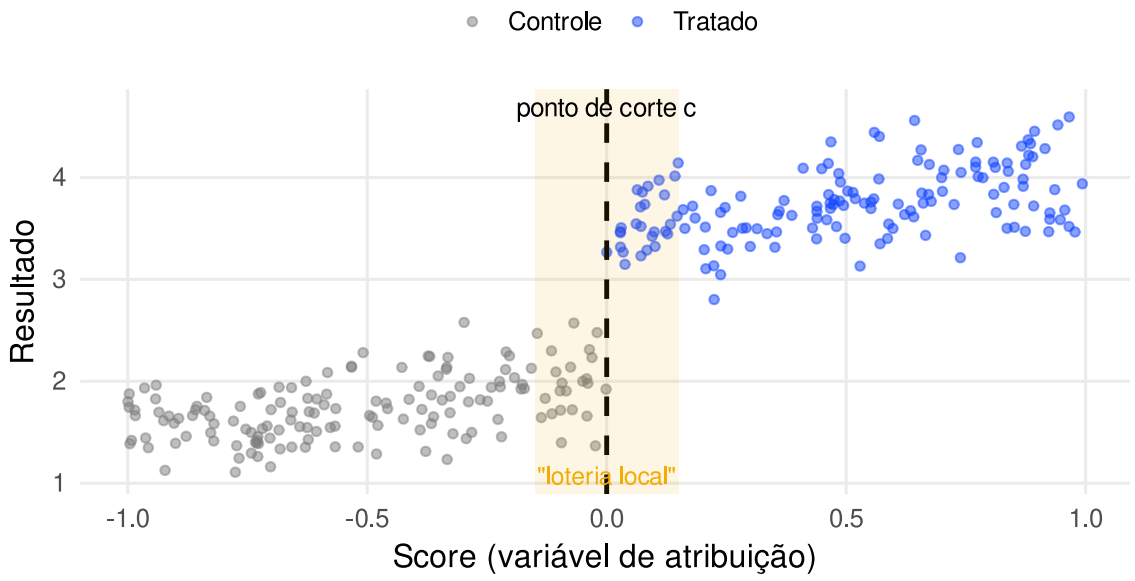
0.4 Regressão discontínua: RDD

- Queremos estimar o **efeito causal** de um tratamento sobre um determinado resultado.
- Mas não podemos simplesmente comparar dois grupos (tratados e não tratados) porque existe auto-seleção, o que gera **viés de seleção**.
- E se a atribuição ao tratamento fosse imposta por uma regra externa? Por exemplo, uma política pública divide a amostra em grupos de tratamento e controle baseando-se em uma **variável contínua** e um determinado valor **limite** (*threshold*)?
- Muitas políticas públicas — bolsas, subsídios, elegibilidade a programas — são **concedidas com base em um critério numérico**. Exemplos:
  - Aluno aprovado para bolsa se nota ≥ 60 pontos
  - Município recebe recursos se IDH < 0,5
  - Candidato eleito se obtém maioria dos votos
- O método de regressão discontínua (RDD) formaliza esse cenário e, sob certas hipótese, **identifica o efeito causal médio local (LATE)**.

0.5 Ponto de corte como “loteria local”

Intuição RDD

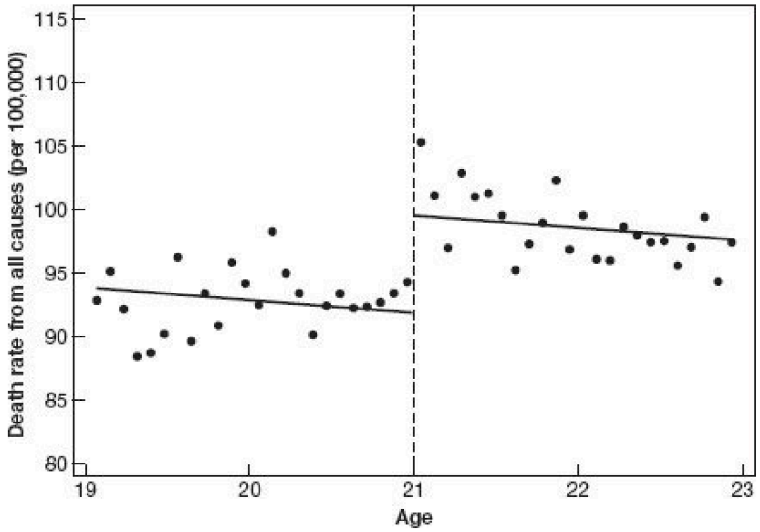
Unidades *muito próximas* ao ponto de corte são quase idênticas: umas recebem o tratamento, outras não, **quase por acaso**.



0.6 Um gráfico vale mais que mil palavras

FIGURE 4.2

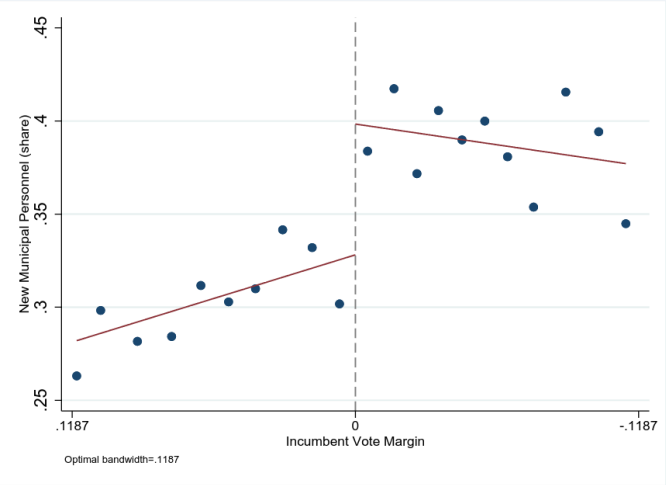
A sharp RD estimate of MLDA mortality effects



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).

0.7 Um gráfico vale mais que mil palavras

Figure 3. : Political Turnover and New Municipal Personnel (Share)



Note: This figure shows the mean of municipal-level share of new municipal personnel in 2009 and 2013 by bins of *IncumbVoteMargin*. Municipalities with *IncumbVoteMargin*<0 experienced change in the political party of the mayor in 2008 or 2012 elections. Municipalities with *IncumbVoteMargin*>0 did not experience change in the political party of the mayor in 2008 or 2012 elections. Note that values to the right side of the zero are negative (political turnover), while values on the left side are positive (no political turnover). New Municipal Personnel (share) is the ratio between the number of new labor contracts that had been added 12 months after the election divided by the total number of contracts that existed the month prior to the election.

0.8 Terminologia

- **Variável de atribuição (*running variable*):** variável/índice contínuo que uma instituição/política/regra usa para **atribuir tratamento** às unidades.
- **Tendências:** referem-se a **mudanças graduais nos resultados** ao longo da variável de atribuição. *OBS:* diferente de tendências temporais
- **Ponto de corte (*cutoff*/*threshold*):** valor específico da variável de atribuição a partir do qual a instituição/política/regra **atribui o tratamento**.
- **Descontinuidade e/ou salto:** como estimamos quebras na trajetória da variável de resultado à direita do ponto de corte, dizemos que existe **descontinuidade** quando isso ocorre.
- **Regressão:** estimações típicas incluem diferença simples de médias, regressões locais por MQO ou regressões globais por MQO em torno do ponto de corte.

0.9 Tipos comuns de RDD

- **Arredondamento:** quando notas em avaliações são contínuas e os valores divulgados são arredondados (ex: Yelp arredonda as avaliações para a **meia estrela mais próxima**, embora a nota subjacente seja contínua).
- **Eleições acirradas:** Votação majoritária, com ponto de corte em 50% dos votos válidos, atribui o poder a diferentes partidos e tipos de candidatos.
- **Algoritmos:** Preços baseados em algorítimos (ex. UBER) são definidos com base em medidas contínuas de demanda, criando limiares onde o preço "salta".

- **Biologia:** Idade mínima para beber legalmente (21 anos nos EUA); peso ao nascer determinando internação em UTI neonatal, idade mínima para aposentadoria.

0.10 Requisitos de dados em RDD

- Amostras **grandes** são uma característica essencial do método de regressão descontínua.
- Se houver tendências significativas na variável de atribuição, é necessário **ainda mais dados** para estimar o salto com precisão.
- Quando há muito ruído nas observações (muita variância) perto do valor de corte, a necessidade de dados aumenta para obter estimativas precisas.
- É preciso ter um número suficiente de observações em torno do ponto de corte para realização de testes de hipótese.
- O método favorece quem tem acesso a dados em nível de firma ou dados administrativos, pois esses bancos de dados tendem a ter muitas observações. (talvez por isso o método se popularizou apenas recentemente)

0.11 Resultados potenciais

Lembre das definições sobre Y, Y^0 e Y^1:

1. Os resultados potenciais têm sobrescritos: Y_i^1, Y_i^0.
   Eles se referem aos resultados hipotéticos de um indivíduo i nos mundos com ou sem tratamento, **independentemente** do que realmente aconteceu.
2. Um desses dois resultados é observado quando o tratamento D_i é definido, de acordo com a equação:
$$Y_i = D_i Y_i^1 + (1 - D_i) Y_i^0$$
3. Os **resultados observados** não têm sobrescritos (Y_i): representam o resultado de fato observado, selecionado entre os dois potenciais, no caso de um tratamento binário.

0.12 Por que isso é importante?

- No método RDD as **hipóteses de identificação** são expressas em termos de $E[Y^1]$ e $E[Y^0]$.
- Na prática, observamos mudanças em $E[Y]$ (não confundir).
- Lembre-se: passamos de resultados potenciais para resultados realizados com base na **atribuição do tratamento**.

0.13 Sharp vs. Fuzzy RDD

Existem dois tipos RDD:

1. **Sharp RDD (nítido):** O tratamento é uma **função determinística da variável de atribuição** (X).

- Exemplo: quem é eleito em uma eleição majoritária (metade dos votos +1).

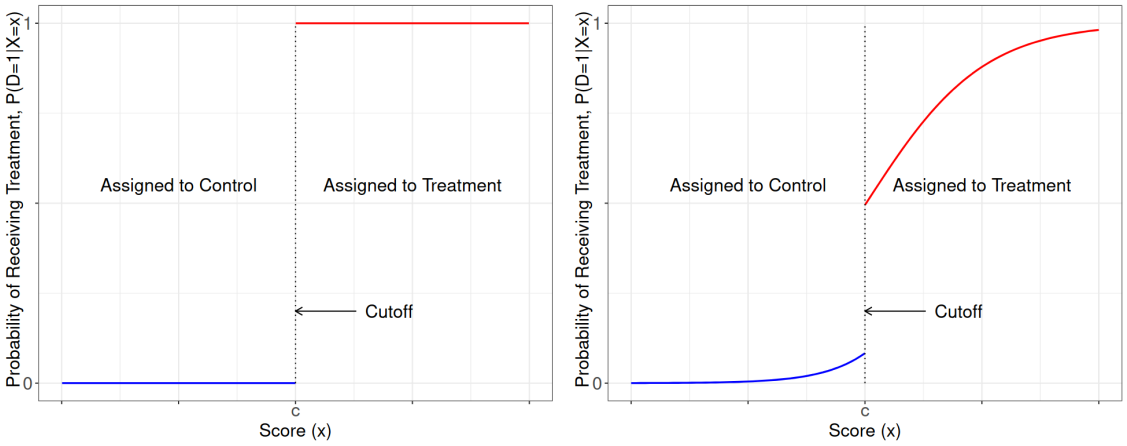
2. **Fuzzy RDD (difuso):** Há uma **descontinuidade na probabilidade de tratamento** quando $X > c_0$. Neste caso, o ponto de corte funciona como um **instrumento (IV)** para o tratamento. (próxima aula)

. . .

Observações:

- O desenho **fuzzy** é um tipo de **estratégia de variáveis instrumentais (IV)** e requer estimadores de **MQ2E**.
- O **sharp**, por outro lado, é uma forma **reduzida** de IV e não requer estimadores **MQ2E**.

0.14 Sharp *vs* Fuzzy RDD



(a) Sharp RD (b) Fuzzy RD

Figure 3.1: Conditional Probability of Receiving Treatment in Sharp vs. Fuzzy RD Designs

0.15 Questões comuns em RDD

- **Suporte comum:** Não há unidades tratadas e de controle ao longo de toda a variável de atribuição, o que torna comparações diretas impossíveis.
- **Extrapolação:** Sem suporte comum, precisamos extrapolar usando modelos — como regressões ou métodos não paramétricos — comparando unidades logo abaixo e acima do ponto de corte.
  - Porém, isso é sensível a tendências, *bandwidths* e ao número de observações.
- **Efeitos do tratamento:** Estimamos efeitos médios do tratamento, mas apenas **para as pessoas no ponto de corte**. Logo, com RDD estimamos o LATE!
  - Resultados podem não ser representativos longe do ponto de corte, especialmente em contextos de alta heterogeneidade.

0.16 Atribuição: RDD Sharp

Atribuição determinística do tratamento

No desenho **Sharp**, o status de tratamento é uma **função determinística e descontínua** de uma variável X_i:

$$D_i = \begin{cases} 1, \text{se} & X_i \geq c_0 \\ 0, \text{se} & X_i < c_0 \end{cases}$$

onde c_0 é um **ponto de corte conhecido**.

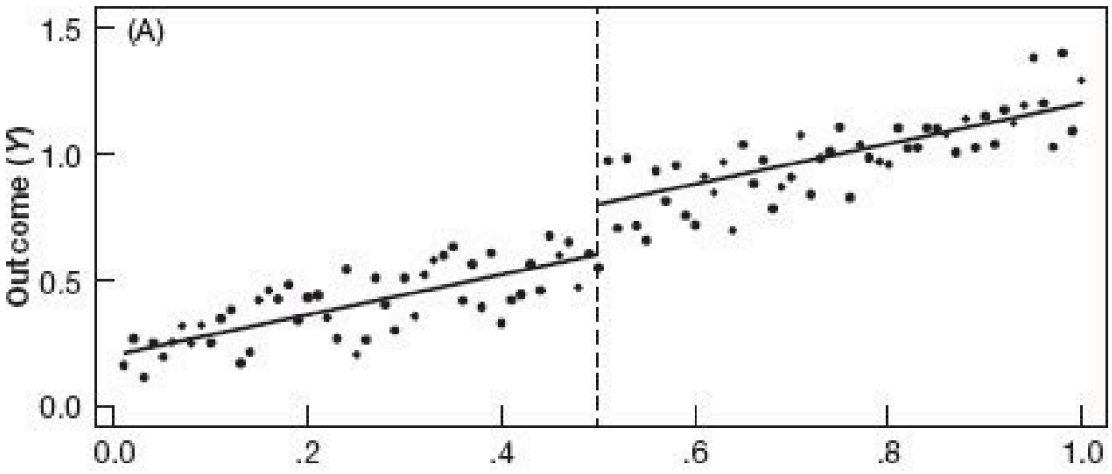
Em outras palavras, se conhecemos o valor de X_i para uma unidade i, sabemos **com certeza** se ela recebeu o tratamento ou não.

0.17 Extrapolação, suporte comum e forma funcional

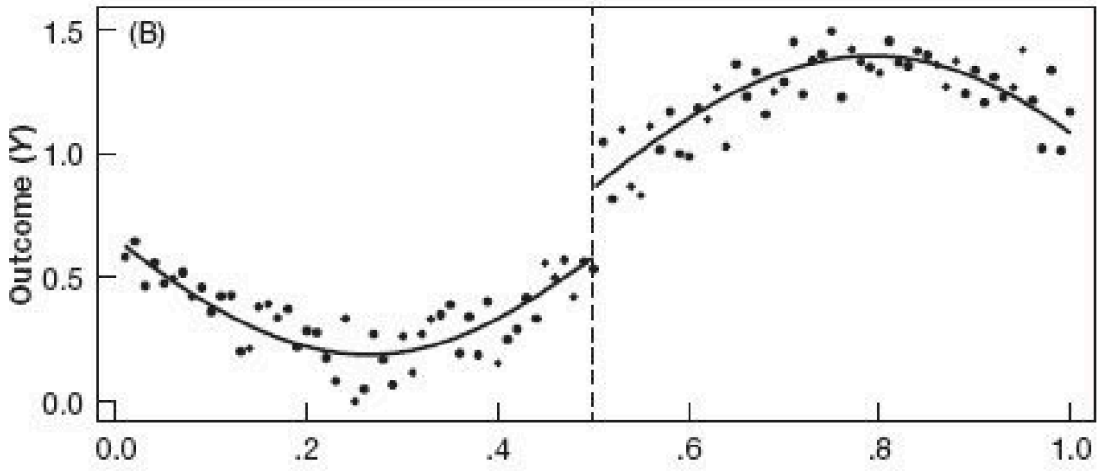
- RDDs Sharp criam problemas de suporte comum, pois **nunca** haverá unidades tratadas e não tratadas ao mesmo tempo ao longo da variável de atribuição.
- Isso exige **extrapolação**: prever resultados fora do suporte dos dados (ou seja, onde o tratamento muda de status no ponto de corte).

- Por isso, as escolhas de modelagem, especialmente a **forma funcional**, são cruciais e representam uma **hipótese crucial**.

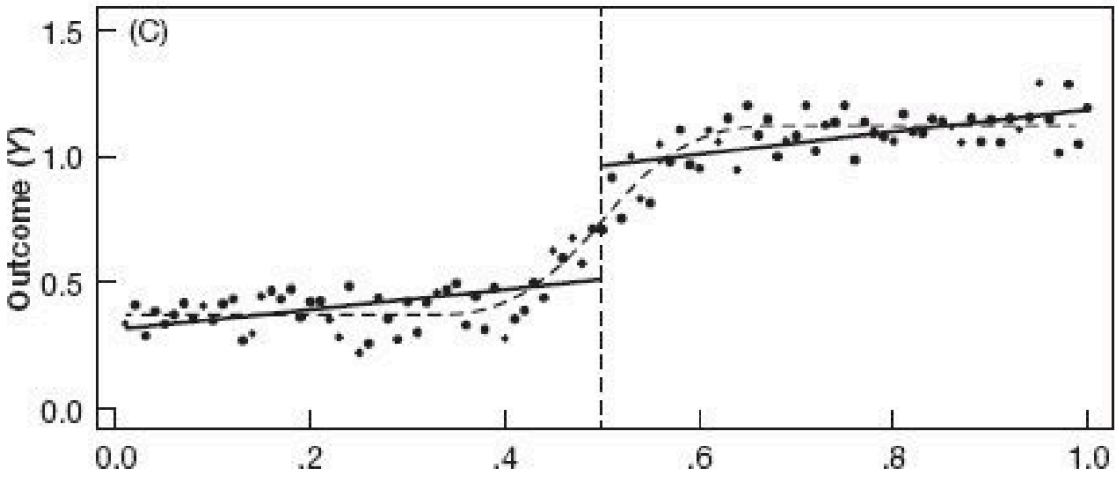
0.18 Modelo Linear com discontinuidade



0.19 Modelo não-linear com discontinuidade



0.20 Não linearidade confundida com descontinuidade



0.21 Efeito do tratamento: LATE

Definição do efeito do tratamento:

O parâmetro de efeito do tratamento δ é a **descontinuidade** na função de expectativa condicional:

$$\begin{aligned}\delta &= \lim_{X_i \to c_0^+} E[Y_i^1 | X_i = c_0] - \lim_{X_i \to c_0^-} E[Y_i^0 | X_i = c_0] \\ &= \lim_{X_i \to c_0^+} E[Y_i | X_i = c_0] - \lim_{X_i \to c_0^-} E[Y_i | X_i = c_0]\end{aligned}$$

A estimação do **Sharp RDD** é interpretada como um **efeito causal médio local (LATE)** do tratamento (D) no ponto de descontinuidade (c_0):

$$\delta_{SRD} = E[Y_i^1 - Y_i^0 \mid X_i = c_0]$$

0.22 Revisão: hipóteses de identificação

1. Hipóteses de independência: típicas em situações com randomização, elimina o viés de seleção.

- Experimentos aleatórios (RCTs)
- Métodos de pareamento e MQO com seleção em observáveis (controles)
- Variáveis instrumentais (IV)

2. Restrições sobre resultados potenciais: associadas a DD e RDD

- Diferenças-em-Diferenças impõe que $E[Y^0|D=1]$ mude ao longo do tempo de modo restrito (tendências paralelas).
- RDD impõe continuidade em $E[Y^0]$ e $E[Y^1]$ em torno do ponto de corte.

0.23 Hipótese de identificação: suavidade

Suavidade das funções de expectativa condicional dos **resultados potenciais**:

$$E[Y_i^0|X = c_0] \text{e} E[Y_i^1|X = c_0]$$

são **contínuas (suaves)** em X no ponto c_0.

- Se as médias populacionais dos **resultados potenciais** $E[Y^1]$ e $E[Y^0]$ forem funções **suaves** de X ao redor do corte c_0, então **não haverá salto** esperado em c_0.
- Isso implica que outros determinantes que poderiam gerar viés (*cofounders*) também evoluem suavemente em torno do ponto de corte.

0.24 Suavidade vs. Efeito do Tratamento

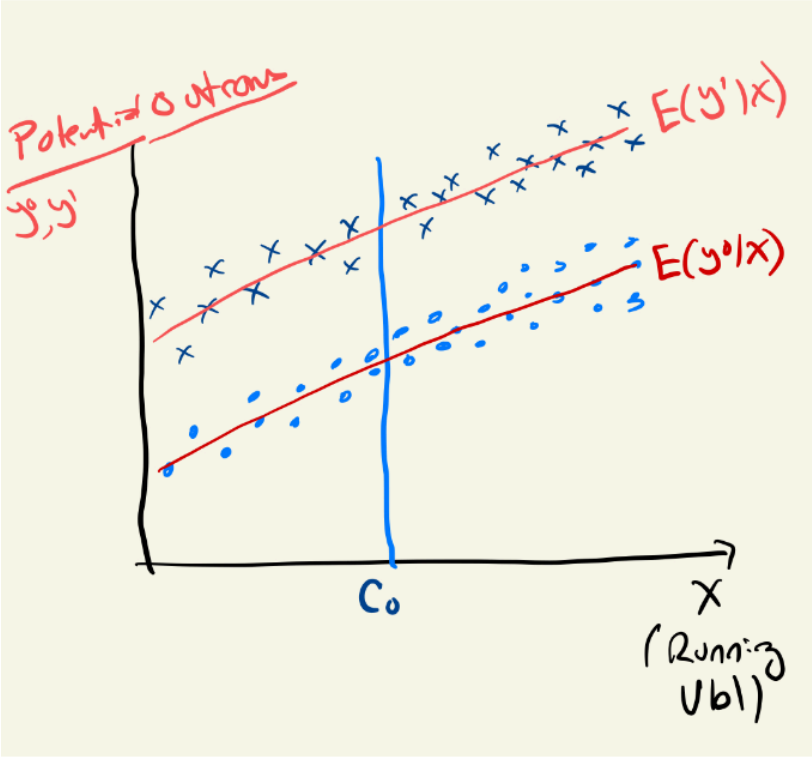


Figura 1: Suavidade dos resultados potenciais

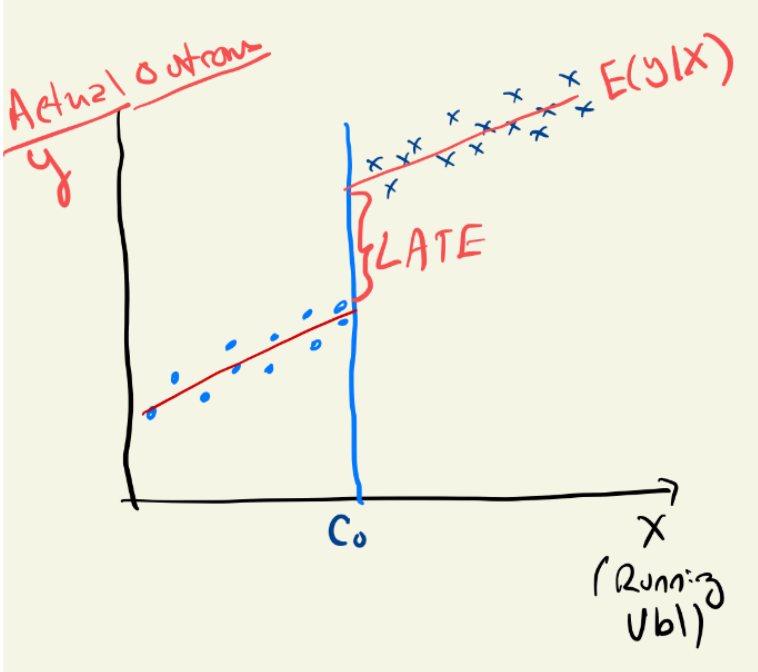


Figura 2: Estimativa do LATE no ponto de corte

Discussão: Por que a figura da esquerda difere da da direita? Para onde foram as duas linhas?

0.25 Resultados potenciais e observados

- **Suavidade** se refere aos **resultados potenciais**:
  - Eles mudam suavemente em média ao longo do ponto de corte.
  - Mas, ao cruzar o ponto de corte, passamos de um conjunto de resultados potenciais a outro.
- **Descontinuidade** se refere aos **resultados realizados**:
  - O ponto de corte **é o mecanismo de atribuição**.
  - Ele alterna os resultados potenciais antes e depois do ponto de corte.
  - Assim, se há um **efeito do tratamento**, observamos um **salto** nos resultados realizados mas requer **extrapolação** para estimação.

0.26 Suavidade permite extrapolação

- A suavidade **justifica o uso de regressões** para extrapolar resultados potenciais faltantes de um lado do corte para o outro
- O **efeito causal médio** é definido **no ponto de corte**, mas a estimação usa dados **à esquerda e à direita** em torno do ponto de corte.
- Lembre-se: **identificação** e **estimação** não são a mesma coisa! É a **suavidade** que permite dar **interpretação causal** às estimativas.
- Estimação se refere a como vamos obter o valor do LATE

0.27 Aproximando a forma funcional

Duas maneiras de estimar o efeito do tratamento em $X = c_0$:

1. Usar **regressões**, com $f(X_i)$ sendo um polinômio de ordem p (altamente sensível à forma funcional).
2. Usar **métodos não-paramétricos**, como **regressões lineares locais** ou **kernels** (menos sensíveis à forma funcional escolhida).

. . .

> ⚠ Aviso
>
> Não iremos cobrir (2) nesse curso pois métodos não-paramétricos não fazem parte da ementa.

0.28 Estimação com extrapolação

- Utilizamos extrapolação para estimar o **efeito médio causal do tratamento** no *Sharp RDD*. A estimativa é não-viesada sob a hipótese de **suavidade**.
- Os modelos predizem contrafactuais esperados usando dados do **lado oposto do ponto de corte**.

0.29 Centralizando a variável de atribuição

Suponha a seguinte função linear:

$$Y_i = \alpha + \beta X_i + \delta D_i + \varepsilon_i$$

Frequentemente centralizamos subtraindo c_0 de X_i:

$$Y_i = \alpha + \beta(X_i - c_0) + \delta D_i + \varepsilon_i$$

- Essa transformação **não altera a interpretação** do efeito do tratamento (δ), apenas modifica o intercepto da regressão.

0.30 Exemplo: Medicare e idade 65

Centralizar a variável de atribuição (idade) subtraindo 65:

$$\begin{aligned} Y &= \beta_0 + \beta_1(\text{Idade}) + \beta_2\ \text{Edu} + \varepsilon \\ &= \beta_0 + \beta_1(\text{Idade} - 65) + \beta_2\ \text{Edu} + \varepsilon \\ &= \beta_0 + \beta_1\ \text{Idade} - \beta_1 65 + \beta_2\ \text{Edu} + \varepsilon \\ &= \alpha + \beta_1\ \text{Idade} + \beta_2\ \text{Edu} + \varepsilon \end{aligned}$$

onde $\alpha = \beta_0 - 65\beta_1$. Quando centralizamos o ponto de corte passa a ser em zero.

> **i** Nota
>
> Todos os demais coeficientes mantêm a **mesma interpretação** apenas o intercepto muda.

0.31 Suavidade vs linearidade

A hipótese de suavidade diz respeito ao comportamento dos **resultados potenciais médios condicionais** quando atravessamos a variável de atribuição e o ponto de corte:

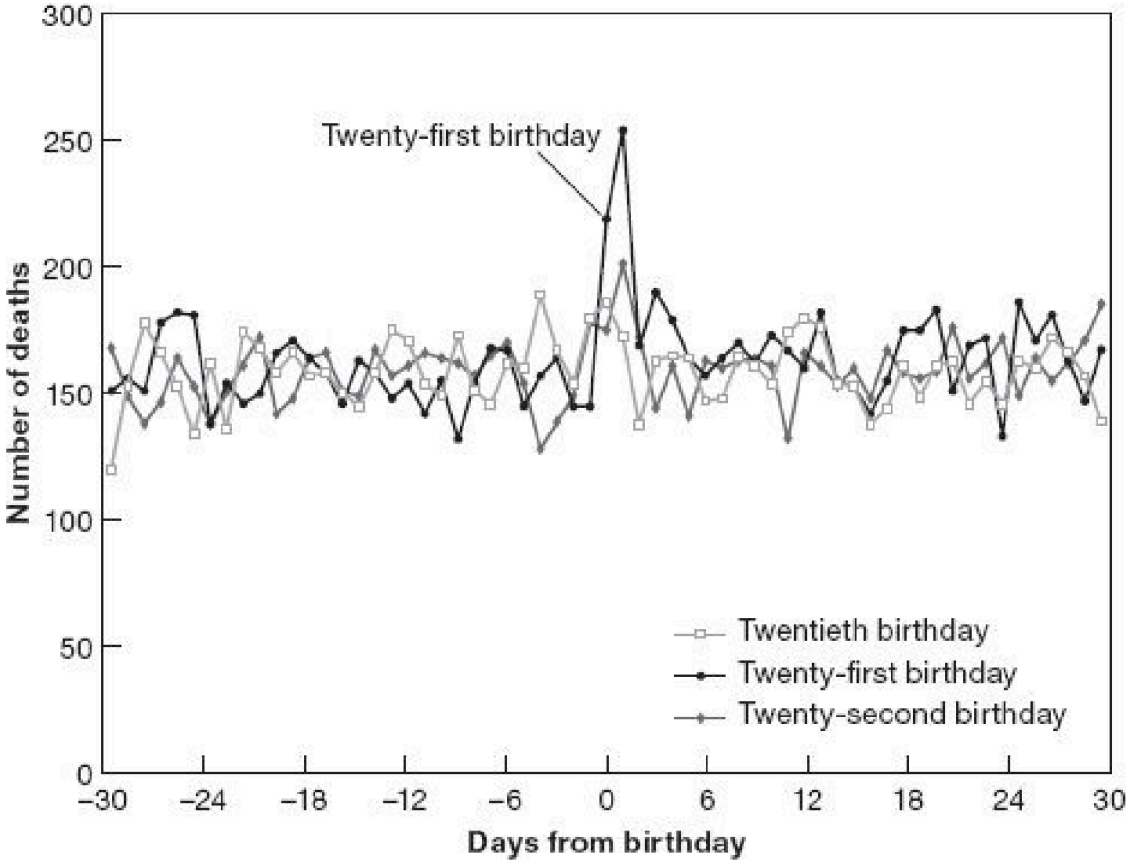
- Isso **não implica linearidade**: a relação pode ser suave **e não linear**.
- Se $E[Y_i^0|X_i]$ não apresenta salto em c_0, mas é **não linear**, podemos obter **resultados espúrios** ao assumir uma forma linear incorreta.
- Suponha que a relação não linear seja $E[Y_i^0|X_i] = f(X_i)$ para alguma função suave $f(X_i)$ (por exemplo, quadrática em X_i).
- Nesse caso, estimamos:$Y_i = f(X_i) + \delta D_i + \eta_i$
- A função $f(X_i)$ **modela os valores contrafactuais** de Y^0. Como estamos extrapolando, precisamos de um **estimador apropriado** que capture bem essa extrapolação.
- Por isso, é comum empregar **transformações polinomiais de ordem mais alta** da variável de atribuição.

1 Idade mínima legal para beber e mortalidade

1.1 Bebida e mortalidade

FIGURE 4.1

Birthdays and funerals



- Questão: **beber causa mais mortes entre jovens adultos?**
- Evidência: nos EUA a **idade legal mínima para beber é 21 anos**.
- Isso cria uma variável de corte **exógena**: antes e depois do aniversário de 21.
- Ao comparar pessoas logo abaixo e acima desse limite, obtemos um **quase-experimento**.
- Ideia: se nada mais muda de forma descontínua aos 21 anos, qualquer salto em mortalidade pode ser atribuído ao consumo de álcool.

1.2 RDD

Variável de atribuição (running variable): idade (X_i)

Cutoff: $c = 21$

Tratamento: $D_i = 1$ se $X_i \geq 21$, 0 caso contrário

Variável de resultado: mortalidade (geral, por acidentes, etc.)

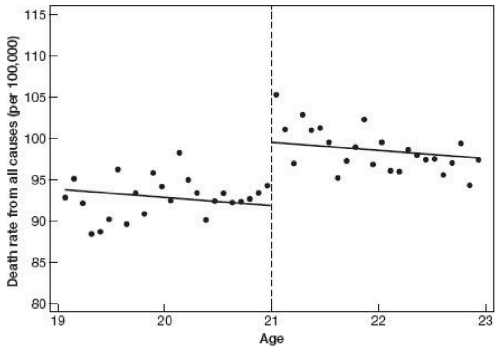
$$Y_i = \alpha + \tau D_i + f(X_i - c) + \varepsilon_i$$

O **parâmetro de interesse τ** é a **mudança discreta** em Y no ponto de corte:

$$\tau = \lim_{x \downarrow c} E[Y_i | X_i = x] - \lim_{x \uparrow c} E[Y_i | X_i = x]$$

1.3 Evidência visual e interpretação

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).

Grande salto em **mortes por acidentes de carro** no aniversário de 21 anos.

Interpretação: aumento repentino no consumo de álcool leva a mais mortes.

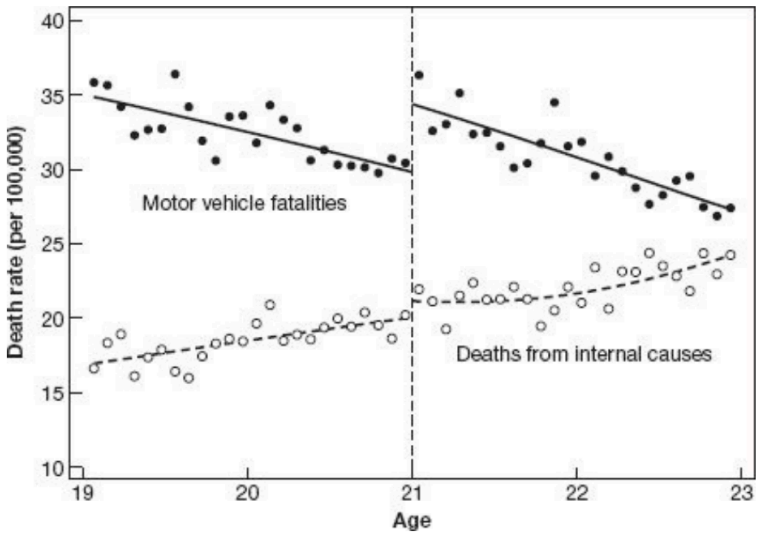
A relação causal é identificada **localmente** (ao redor do cutoff).

RD fornece um **efeito causal local (LATE)** em torno do limiar.

1.4 Resultados

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).

1.5 Resultados

TABLE 4.1

Sharp RD estimates of MLDA effects on mortality

Dependent variable	Ages 19–22		Ages 20–21	
	(1)	(2)	(3)	(4)
All deaths	7.66	9.55	9.75	9.61
	(1.51)	(1.83)	(2.06)	(2.29)
Motor vehicle accidents	4.53	4.66	4.76	5.89
	(.72)	(1.09)	(1.08)	(1.33)
Suicide	1.79	1.81	1.72	1.30
	(.50)	(.78)	(.73)	(1.14)
Homicide	.10	.20	.16	−.45
	(.45)	(.50)	(.59)	(.93)
Other external causes	.84	1.80	1.41	1.63
	(.42)	(.56)	(.59)	(.75)
All internal causes	.39	1.07	1.69	1.25
	(.54)	(.80)	(.74)	(1.01)
Alcohol-related causes	.44	.80	.74	1.03
	(.21)	(.32)	(.33)	(.41)
Controls	age	age, age^2, interacted with over-21	age	age, age^2, interacted with over-21
Sample size	48	48	24	24

Notes: This table reports coefficients on an over-21 dummy from regressions of month-of-age-specific death rates by cause on an over-21 dummy and linear or interacted quadratic age

2 Political Turnover, Bureaucratic Turnover and the Quality of Public Services

2.1 Pergunta de Pesquisa

- Moreira, Akhtari e Trucco investigam **como a rotatividade política em eleições municipais afeta a burocracia e a qualidade dos serviços públicos** no Brasil.
- Foco principal: **impacto da mudança de partido do prefeito** sobre a provisão de educação pública municipal.
- Questão central: A substituição política de burocratas após uma eleição melhora ou prejudica a qualidade dos serviços públicos?
- Motivação: Em contextos onde políticos controlam nomeações, **a mudança de governo pode gerar substituições generalizadas** com possíveis efeitos negativos sobre continuidade, conhecimento institucional e desempenho do serviço público.

2.2 Contexto e Motivação

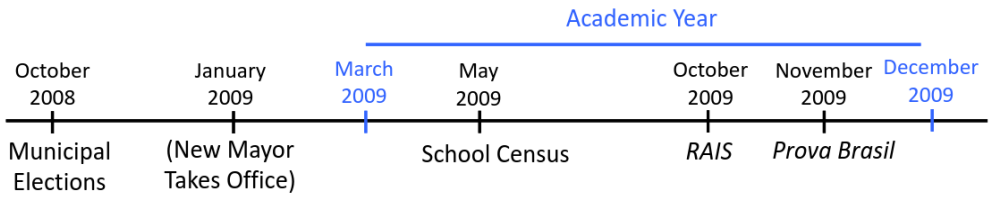
- O estudo analisa **eleições municipais brasileiras**, onde prefeitos controlam fortemente o funcionalismo local, incluindo diretores e professores de escolas municipais.
- O Brasil é um dos países com **maior discricionariedade política** sobre cargos públicos na América Latina.
- A literatura sugere que:
  - Mudanças políticas podem gerar **benefícios políticos de curto prazo** (lealdade, patronagem).
  - Mas também podem implicar **perda de capital humano e disrupção organizacional**.
- A educação é o setor ideal para estudar esse efeito, pois:
  - É **responsabilidade central dos municípios** (≈30% do orçamento).
  - Há **dados ricos sobre desempenho** (testes padronizados e registros administrativos).

2.3 Dados e estratégia de identificação

- **Base de dados combinada:**
  - Resultados eleitorais do TSE (2004–2012).
  - Dados administrativos de pessoal (RAIS).
  - Censo Escolar e Prova Brasil (INEP).
  - Finanças municipais (FINBRA) e dados socioeconômicos (IBGE/IPEA).
- **Estratégia causal:**
  - **Regressão discontínua (RDD)** em eleições municipais.
  - Compara municípios onde o partido incumbente **venceu por pequena margem** (sem troca) com aqueles onde **perdeu por pequena margem** (com troca).
  - Suposição de identificação: Em eleições acirradas, a vitória ou derrota do partido é "quase aleatória".
- Isso permite estimar o **efeito causal da mudança partidária** sobre a rotatividade burocrática e a qualidade da educação (testes de alunos).

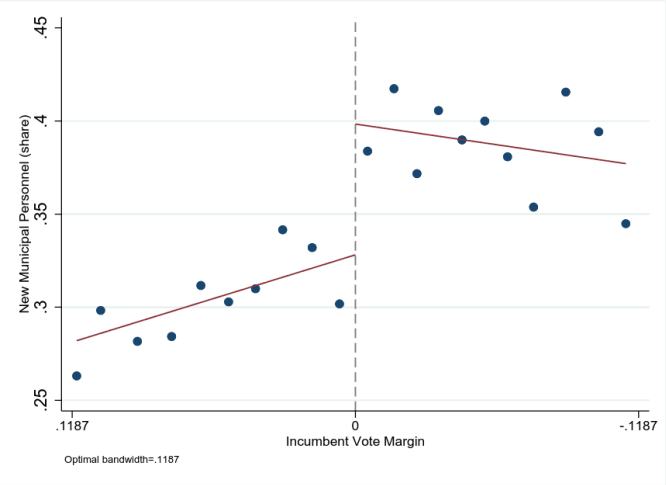
2.4 Timeline e estratégia empírica

Figure 2. : Timeline of Election and Data Collection



2.5 Resultados: contratação

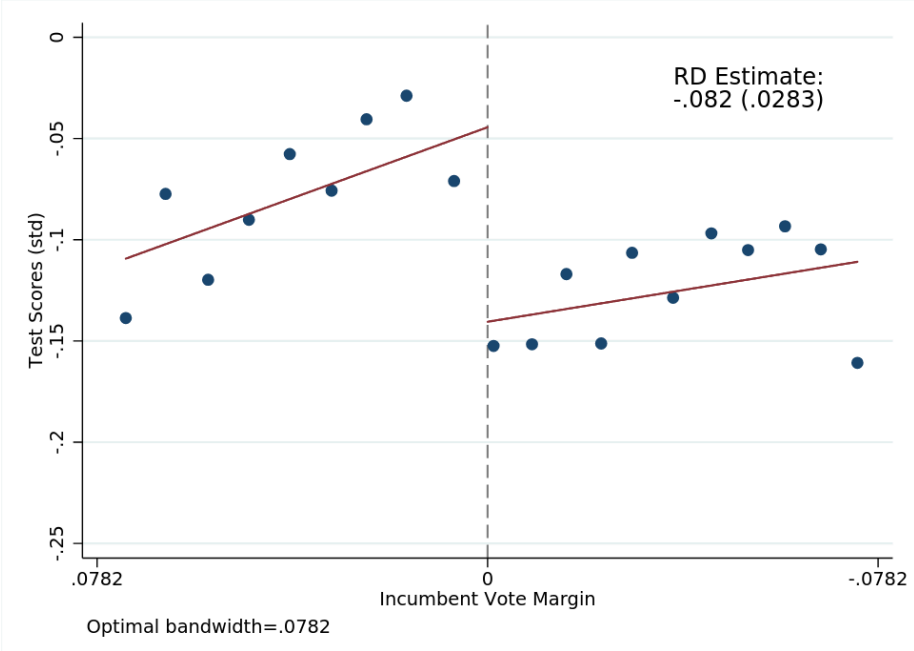
Figure 3. : Political Turnover and New Municipal Personnel (Share)



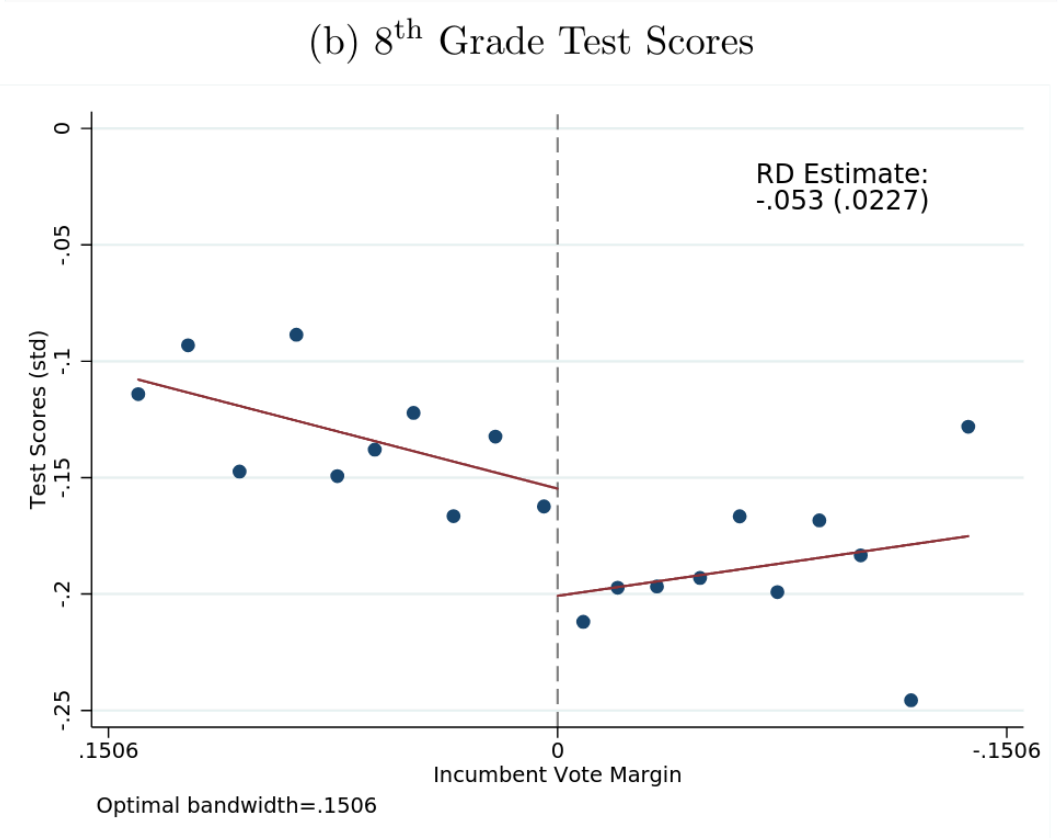
Note: This figure shows the mean of municipal-level share of new municipal personnel in 2009 and 2013 by bins of *IncumbVoteMargin*. Municipalities with *IncumbVoteMargin*<0 experienced change in the political party of the mayor in 2008 or 2012 elections. Municipalities with *IncumbVoteMargin*>0 did not experience change in the political party of the mayor in 2008 or 2012 elections. Note that values to the right side of the zero are negative (political turnover), while values on the left side are positive (no political turnover). New Municipal Personnel (share) is the ratio between the number of new labor contracts that had been added 12 months after the election divided by the total number of contracts that existed the month prior to the election.

2.6 Resultados: notas 4a. série

(a) 4^{th} Grade Test scores



2.7 Resultados: notas 8a. série



2.8 Resultados: diretores

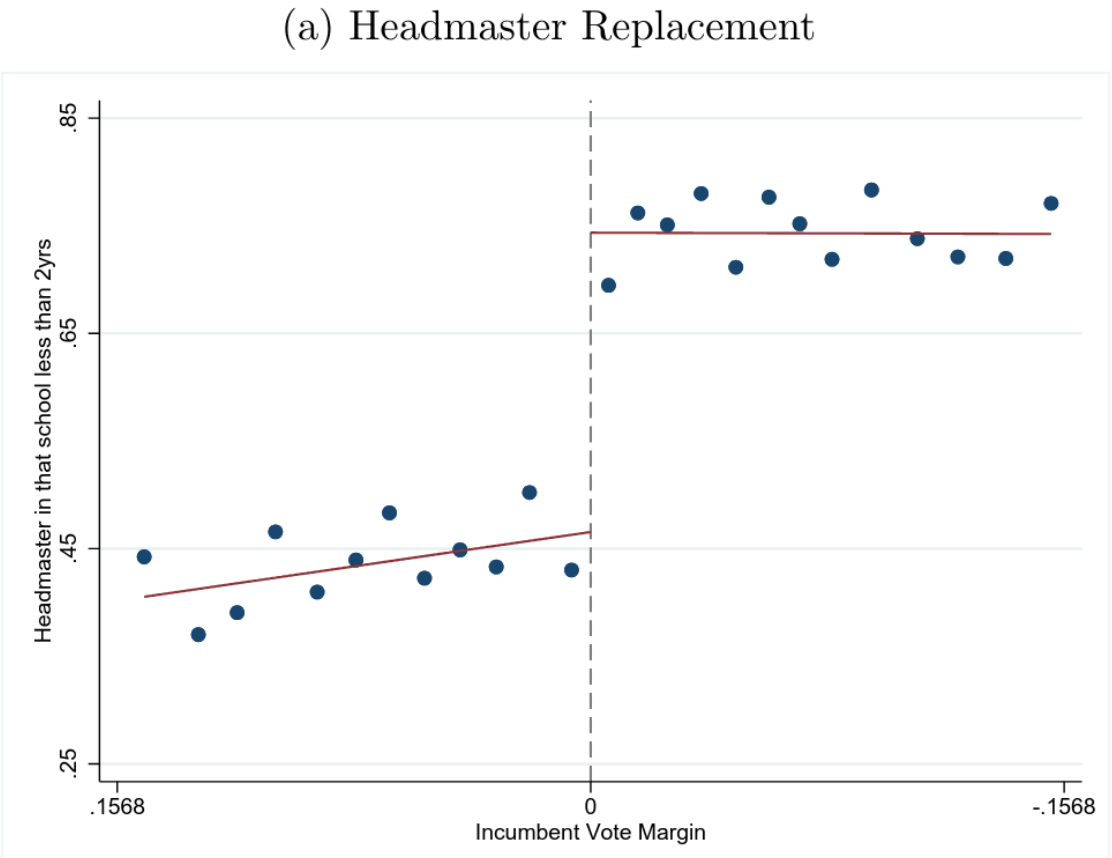


Figura 3: Escolas municipais

(a) Headmaster Replacement

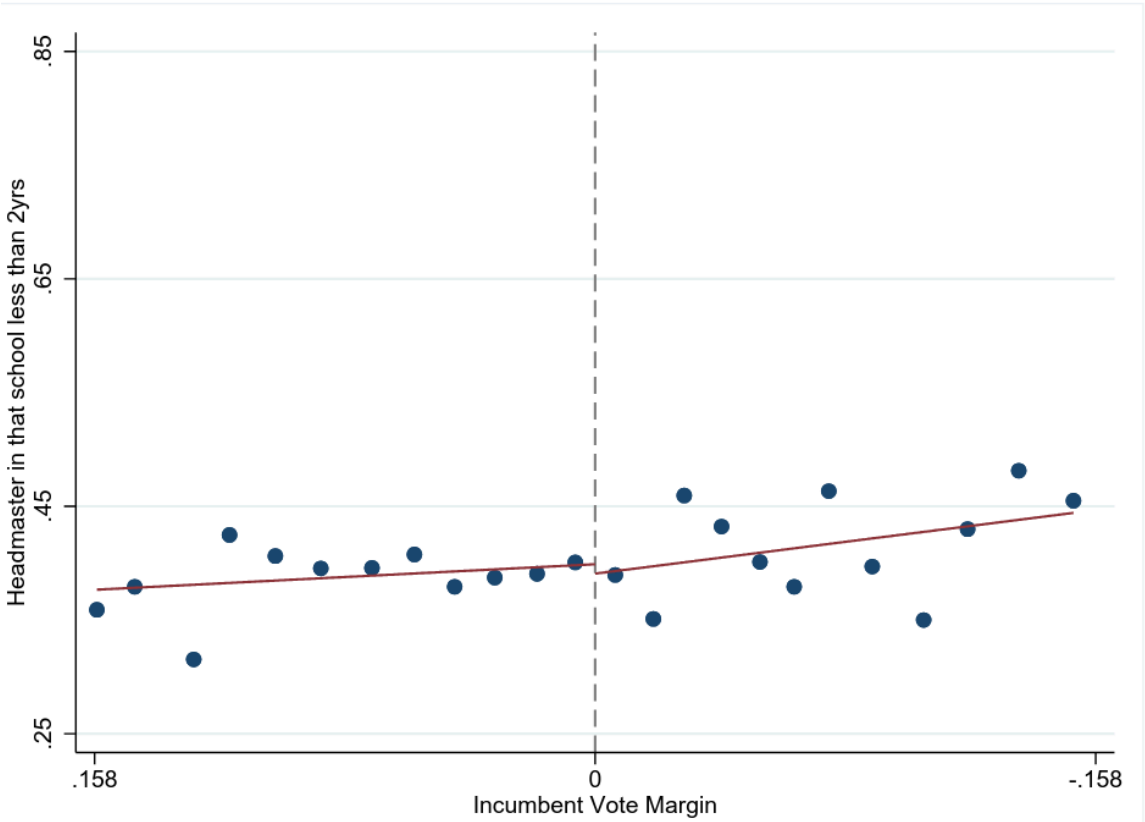


Figura 4: Escolas não municipais

2.9 Resultados: profesores

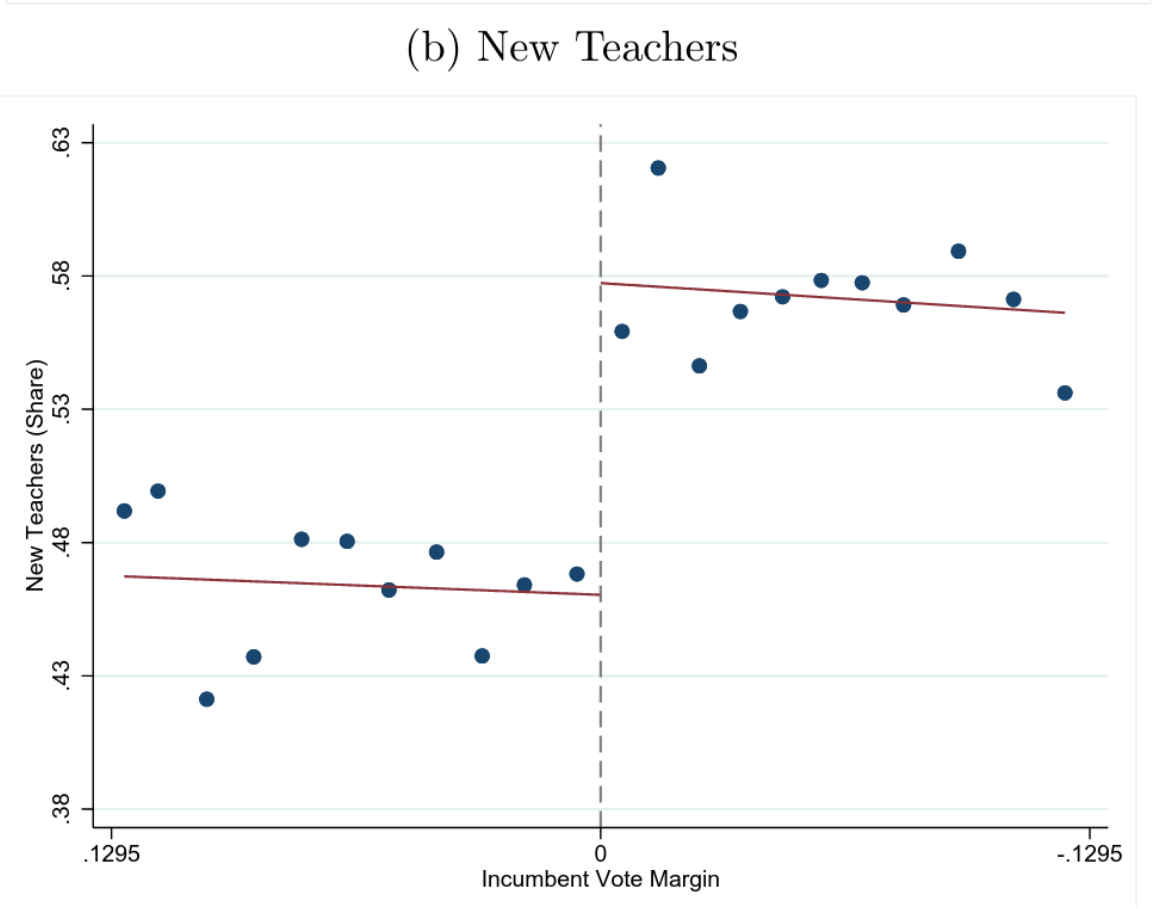


Figura 5: Escolas municipais

(b) New Teachers

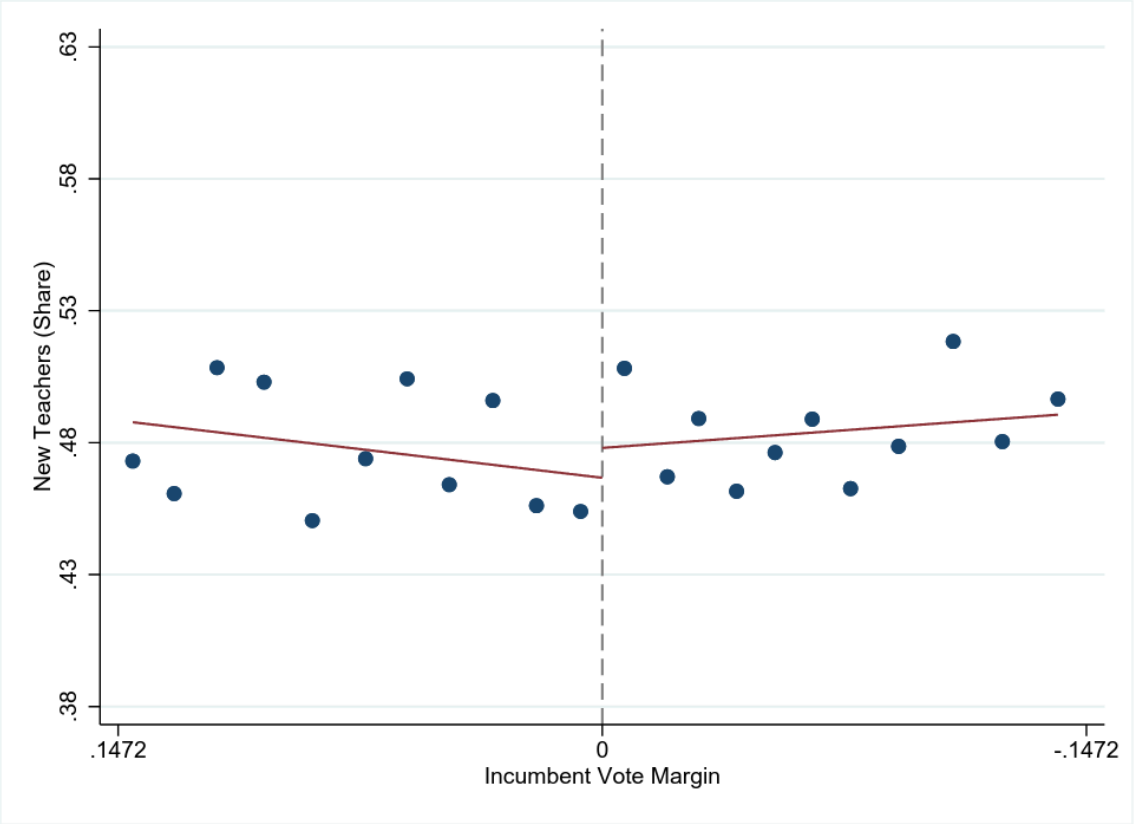


Figura 6: Escolas não municipais