Regressão Descontinuada

Inferência Causal (MI628A)

Marília Rocha
Thiago Paulichen
Tiago Amorim
IMECC - Unicamp

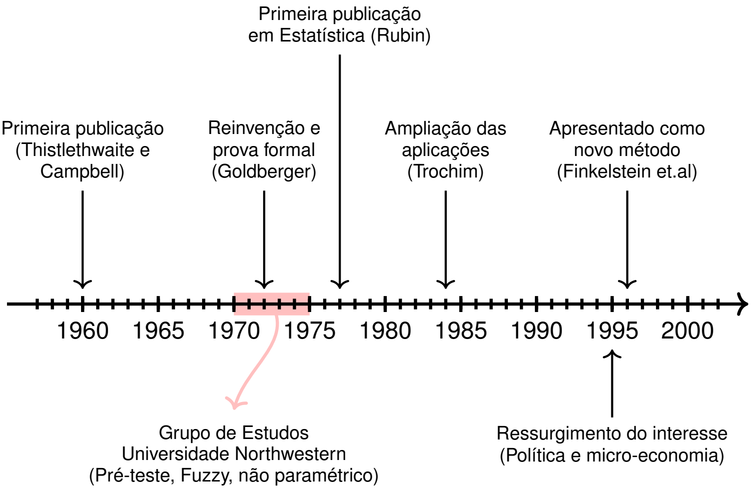
25 de Junho de 2024

A Regressão Descontinuada (*Regression Discontinuity Design* - RDD) é utilizada quando a atribuição do tratamento depende deterministicamente de uma covariável. Quando essa atribuição é exata, o processo de seleção é totalmente conhecido e pode ser modelado para produzir uma inferência causal não-viesada.

Adaptado de: *Waiting for Life to Arrive*
THOMAS D COOK

Sumário

Linha do Tempo



Levantamento de Cook [2008]

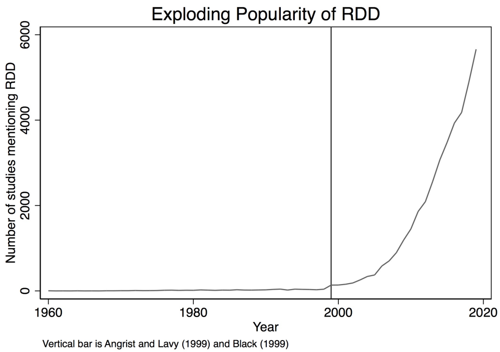
Dificuldades e Ressurgimento

Impasses:

- Preconceito da comunidade estatística por um tema desenvolvido pelas ciências sociais;
- Desenvolvimento inicial restrito ao grupo da Northwestern;
- Papers usavam diferentes termos para RDD.

Razões para o ressurgimento:

- Desenvolvimento por economistas renomados em diversas instituições;
- Nova gama de aplicações.



Fonte: Cunningham [2021]

Maiores detalhes em Cook [2008].

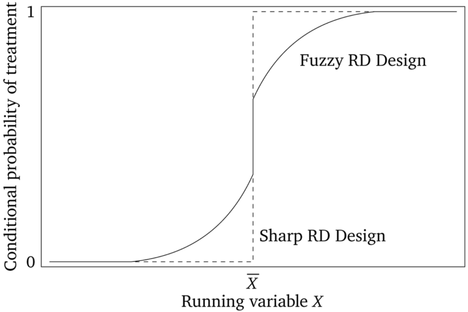
Descontinuidade *Sharp* e *Fuzzy*

Sharp: a atribuição de tratamento segue uma regra determinista:

$$Z_i = \begin{cases} 1 & \text{se } x_i \geq c \\ 0 & \text{c.c.} \end{cases}$$

Fuzzy: a probabilidade de atribuição de tratamento é descontínua em um ponto de corte conhecido:

$$\mathbb{P}(Z_i = 1 \mid x_i) = \begin{cases} g_1(x_i) & \text{se } x_i \geq c \\ g_0(x_i) & \text{c.c.} \end{cases}$$



Fonte: Cunningham [2021]

Descontinuidade *Sharp* e *Fuzzy*

Sharp: a atribuição de tratamento segue uma regra determinista:

$$Z_i = \begin{cases} 1 & \text{se } x_i \geq c \\ 0 & \text{c.c.} \end{cases}$$

Fuzzy: a probabilidade de atribuição de tratamento é descontínua em um ponto de corte conhecido:

$$\mathbb{P}(Z_i = 1 \mid x_i) = \begin{cases} g_1(x_i) & \text{se } x_i \geq c \\ g_0(x_i) & \text{c.c.} \end{cases}$$

Exemplos:

- S Droga administrada a pacientes com uma taxa acima de certo limite.
- S Aulas de recuperação obrigatória para alunos com média abaixo de certo valor.
- F Programa da saúde suplementar oferecido à famílias que se enquadram em determinado critério de renda.

Descontinuidade *Sharp* e *Fuzzy*

Sharp RD

Tratamento é determinado como $Z_i = \mathbb{I}[X_i \geq c]$, onde X_i é a variável de atribuição e c é o ponto de corte.

Suposições:

(A1) $\mathbb{E}(Y_i(0) \mid X_i)$ é contínuo em $X_i = c$;

(A2) $\mathbb{E}(Y_i(1) \mid X_i)$ é contínua em $X_i = c$;

(A3) $\mathbb{E}(Y_i(1) - Y_i(0) \mid X_i)$ é contínua em $X_i = c$;

Um design RD foca no ponto de corte na estimação do efeito causal:

$$\tau_c := \mathbb{E}(\tau_i \mid X_i = c) = \mathbb{E}(Y_i(1) - Y_i(0) \mid X_i = c),$$

e considera principalmente amostras localmente próximas ao ponto de corte.

Sharp RD

Dado $\varepsilon > 0$ e as suposições **(A1)**, **(A2)** e **(A3)**, podemos tomar os limites laterais ao longo do ponto de corte c, obtendo:

$$\begin{aligned}\mathbb{E}(Y_i(1) \mid X_i = c) &= \lim_{\varepsilon \to 0} \mathbb{E}(Y_i(1) \mid X_i = c + \varepsilon) \\ &= \lim_{\varepsilon \to 0} \mathbb{E}(Y_i(1) \mid Z_i = 1, X_i = c + \varepsilon) \\ &= \lim_{\varepsilon \to 0} \mathbb{E}(Y_i \mid X_i = c + \varepsilon).\end{aligned}$$

Similarmente:

$$\mathbb{E}(Y_i(0) \mid X_i = c) = \lim_{\varepsilon \to 0} \mathbb{E}(Y_i \mid X_i = c - \varepsilon).$$

Dessa forma temos que:

$$\tau_c = \lim_{\varepsilon \to 0} \left[\mathbb{E}(Y_i \mid X_i = c + \varepsilon) - \mathbb{E}(Y_i \mid X_i = c - \varepsilon)\right].$$

Fuzzy RD

A probabilidade condicional do tratamento $\mathbb{P}(Z_i = 1 \mid X_i)$, não pula de 0 para 1 no ponto de corte. $\mathbb{P}(Z_i = 1 \mid X_i)$ é descontínua em c e o tamanho dessa descontinuidade está entre 0 e 1.

Suposição adicional:

(A4) $Y_i(0), Y_i(1) \perp Z_i \mid X_i$ (ignorabilidade).

Com isso, para um $\varepsilon > 0$ dado, temos:

$$\begin{aligned}
&\mathbb{E}(Y_i \mid X_i = c + \varepsilon) - \mathbb{E}(Y_i \mid X_i = c - \varepsilon) \\
&= \mathbb{E}(Y_i(0) + Z_i\tau_i \mid X_i = c + \varepsilon) - \mathbb{E}(Y_i(0) + Z_i\tau_i \mid X_i = c - \varepsilon) \\
\textbf{(A4)}\ &= \mathbb{E}(Y_i(0) \mid X_i = c + \varepsilon) + \mathbb{E}(Z_i \mid X_i = c + \varepsilon)\mathbb{E}(\tau_i \mid X_i = c + \varepsilon) - \mathbb{E}(Y_i(0) \mid X_i = c - \varepsilon) \\
&\quad - \mathbb{E}(Z_i \mid X_i = c - \varepsilon)\mathbb{E}(\tau_i \mid X_i = c - \varepsilon).
\end{aligned}$$

Fuzzy RD

Fazendo o limite $\varepsilon \to 0$, segue que:

$$\lim_{\varepsilon\to 0}\left[\mathbb{E}(Y_i \mid X_i = c+\varepsilon) - \mathbb{E}(Y_i \mid X_i = c-\varepsilon)\right] = \left(\lim_{\varepsilon\to 0}\left[\mathbb{E}(Z_i \mid X_i = c+\varepsilon) - \mathbb{E}(Z_i \mid X_i = c-\varepsilon)\right]\right)\tau_c$$

isto é:

$$\tau_c = \lim_{\varepsilon\to 0}\left[\frac{\mathbb{E}(Y_i \mid X_i = c+\varepsilon) - \mathbb{E}(Y_i \mid X_i = c-\varepsilon)}{\mathbb{E}(Z_i \mid X_i = c+\varepsilon) - \mathbb{E}(Z_i \mid X_i = c-\varepsilon)}\right].$$

Observe que o caso *Sharp* pode ser visto como um caso particular do *Fuzzy*, uma vez que neste caso:

$$\lim_{\varepsilon\to 0}\left[\mathbb{E}(Z_i \mid X_i = c+\varepsilon) - \mathbb{E}(Z_i \mid X_i = c-\varepsilon)\right] = 1.$$

Estimadores Locais

Estimador local não-paramétrico:

$$\hat{\tau}_c = \frac{\sum_{i:c \leq X_i < c+h} Y_i K\left(\frac{X_i - c}{h}\right)}{\sum_{i:c \leq X_i < c+h} K\left(\frac{X_i - c}{h}\right)} - \frac{\sum_{i:c-h < X_i < c} Y_i K\left(\frac{X_i - c}{h}\right)}{\sum_{i:c-h < X_i < c} K\left(\frac{X_i - c}{h}\right)},$$

onde $K(u)$ é um kernel com $\int_{-1}^{1} K(u)\, du = 1$. É viesado em $\mathcal{O}(h)$.

Estimador local linear (*Sharp* RD):

$$\hat{\tau}_c = \hat{\beta}_0^+ - \hat{\beta}_0^-$$

onde $\hat{\beta}_0^+$ e $\hat{\beta}_0^-$ vem do ajuste de Y_i por mínimos quadrados:

$$\min_{\beta^+} \sum_{i:c \leq X_i < c+h} (Y_i - \beta_0^+ - \beta_x^+ (X_i - c))^2 \quad \text{e} \quad \min_{\beta^-} \sum_{i:c-h < X_i < c} (Y_i - \beta_0^- - \beta_x^- (X_i - c))^2.$$

Viés com regressão é em geral $\mathcal{O}(h^2)$.

Estimadores Locais

Estimador local linear (*Fuzzy* RD):

$$\hat{\tau}_c = \frac{\hat{\tau}_y}{\hat{\tau}_z} = \frac{\hat{\beta}_0^+ - \hat{\beta}_0^-}{\hat{\alpha}_0^+ - \hat{\alpha}_0^-}$$

onde $\hat{\beta}_0^+$ e $\hat{\beta}_0^-$ seguem fórmula para descontinuidade *sharp* e $\hat{\alpha}_0^+$ e $\hat{\alpha}_0^-$ vem do ajuste de Z_i por mínimos quadrados:

$$\min_{\alpha^+} \sum_{i: c \leq X_i < c+h} (Z_i - \alpha_0^+ - \alpha_x^+(X_i - c))^2$$

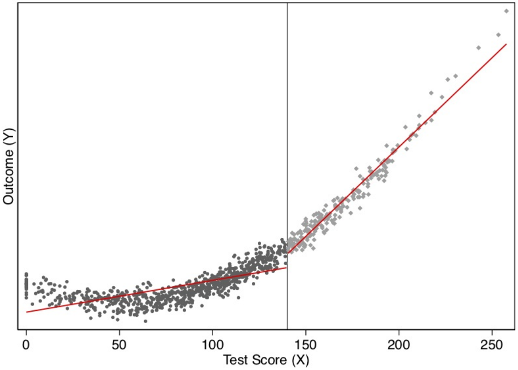
$$\min_{\alpha^-} \sum_{i: c-h < X_i < c} (Z_i - \alpha_0^- - \alpha_x^-(X_i - c))^2.$$

Imbens and Lemieux [2008] sugerem **fortemente** fazer regressão.

É sugerido usar $h \propto N^{-\delta}$, com $1/5 < \delta < 2/5$.

h ótimo estimado com validação cruzada.

É preciso tomar cuidado para não confundir descontinuidade com não-linearidade!



Fonte: Cunningham [2021]

Estimador da Variância

Distribuição assintótica:

$$\sqrt{Nh}(\hat{\tau} - \tau) \to \mathcal{N}\left(0, \frac{1}{\tau_z^2} V_{\tau_y} + \frac{\tau_y^2}{\tau_z^4} V_{\tau_z} - 2\frac{\tau_y}{\tau_z^3} C_{\tau_y \tau_z}\right).$$

Um estimador *pluggin* é estimar os termos da equação acima.

- $\hat{V}_{\tau_y} = \frac{4}{\hat{f}_x(c)}(\hat{\sigma}^2_{y^+} + \hat{\sigma}^2_{y^-})$
- $\hat{V}_{\tau_z} = \frac{4}{\hat{f}_x(c)}(\hat{\sigma}^2_{z^+} + \hat{\sigma}^2_{z^-})$
- $\hat{C}_{\tau_y \tau_z} = \frac{4}{\hat{f}_x(c)}(\hat{C}_{yz^+} + \hat{C}_{yz^-})$
- $\hat{f}_x(x) = \frac{N_{h^+} + N_{h^-}}{2Nh}$
- $\hat{\sigma}^2_{y^+} = \frac{1}{N_{h^+}} \sum_{i: c \leq X_i < c+h} (Y_i - \hat{Y}(X_i))^2$
- $\hat{\sigma}^2_{z^+} = \frac{1}{N_{h^+}} \sum_{i: c \leq X_i < c+h} (Z_i - \hat{Z}(X_i))^2$
- $\hat{C}_{yz^+} = \frac{1}{N_{h^+}} \sum_{i: c \leq X_i < c+h} (Y_i - \hat{Y}(X_i))(Z_i - \hat{Z}(X_i))$

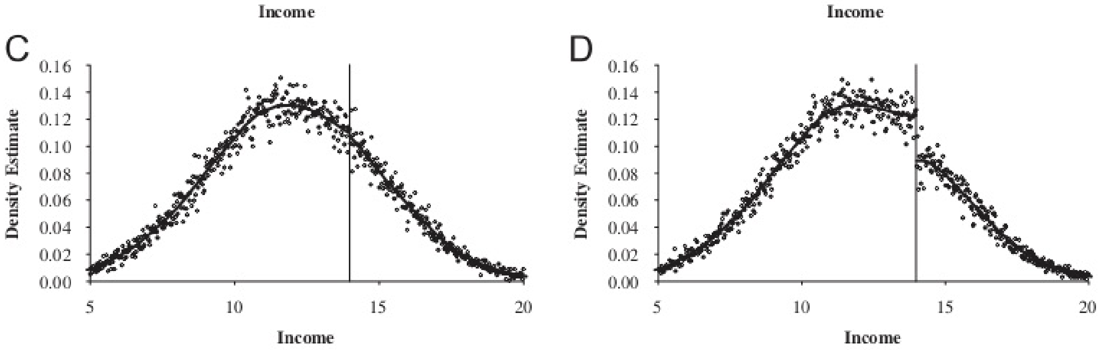
Os termos *negativos* tem somatório em $\{i : c - h < X_i < c\}$, e usam N_{h^-}.

Pode-se substituir pelo estimador robusto de um TSLS para RD *Fuzzy* e de OLS para RD *Sharp*.

Teste de Densidade de McCrary

O teste de densidade de McCrary [2008] ajuda a avaliar a validade dos dados.

Avaliação envolve o uso de polinômios locais para estimar densidade.

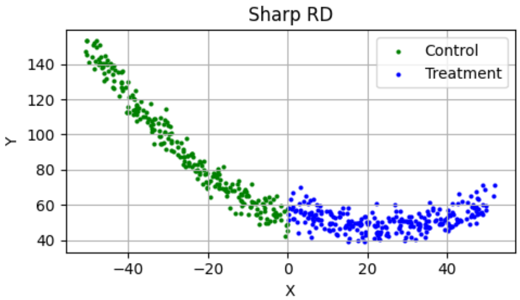


Fonte: Cunningham [2021]

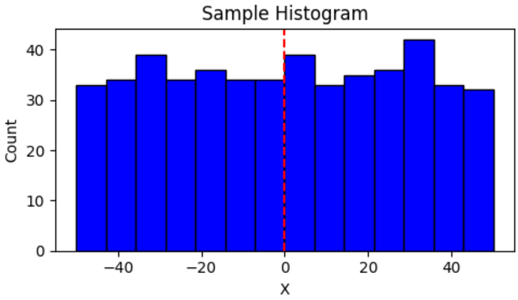
Toy Problem

Função geradora dos dados:

$$Y_i = 50 - X_i + 0.02X_i^2 + \tau\mathbb{I}(X_i > 0) + \mathcal{N}(0, 5)$$



Avaliação da densidade de pontos.

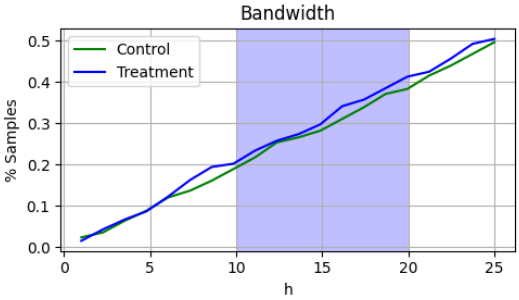


Não foi encontrado código pronto em Python para cálculo da densidade local.

Em R: **rddensity**.

Toy Problem

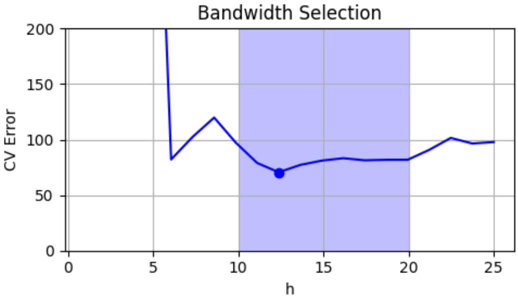
Largura da banda sugerida: h equivalente a 20 a 40% dos pontos de cada lado.



Otimização feita com validação cruzada com 50% dos de cada lado.

$$CV(h) = \frac{1}{N} \sum_i (Y_i - \hat{Y}_h(X_i))^2$$

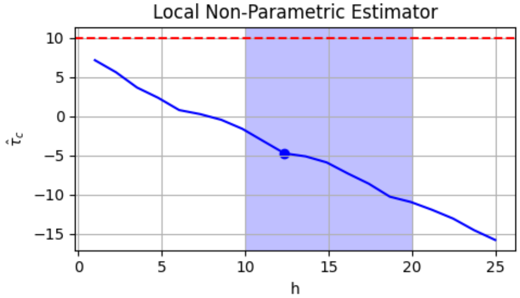
$$h^* = \underset{h}{\operatorname{argmin}}\, CV(h)$$

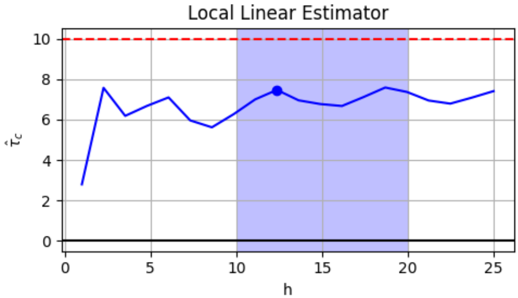


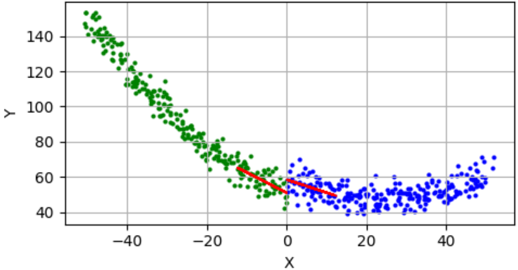
Toy Problem

Estimador não-paramétrico local: retângulo em [c-h,c+h].

$$\hat{\tau}_c = \frac{\sum_{i=1}^{N} Y_i \mathbb{I}(c \leq X_i < c+h)}{\sum_{i=1}^{N} \mathbb{I}(c \leq X_i < c+h)} - \frac{\sum_{i=1}^{N} Y_i \mathbb{I}(c-h < X_i < c)}{\sum_{i=1}^{N} \mathbb{I}(c-h < X_i < c)}.$$



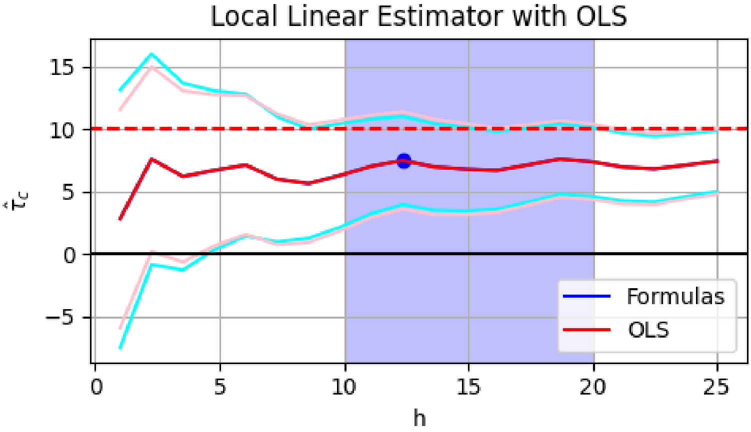




Toy Problem

Para utilizar mínimos quadrados é construída uma aproximação conjunta dos dois *lados*:

$$\min_{\beta} \sum_{i:c-h<X_i<c+h} (Y_i - \beta_0 - \beta_x^+(X_i - c)\mathbb{I}(c \leq X_i < c+h) - \beta_x^-(X_i - c)\mathbb{I}(c-h < X_i < c) - \beta_\tau Z_i)^2$$



Medicare

Objetivo: Avaliar o impacto do plano de saúde na utilização de cuidados médicos. (Card et al. [2008])

Limitações: Heterogeneidade de cobertura, viés de seleção - oferta e procura dependem da saúde inicial, confundindo comparações observacionais.

Solução: Abordagem de regressão descontinuada para comparar o estado de saúde entre pessoas imediatamente antes e imediatamente após os 65 anos de idade (elegibilidade para o programa *Medicare*):

- Mudanças no número de consultas médicas recentes e nas internações hospitalares;
- Efeitos em diferentes subgrupos;
- Quantificar até que ponto o início da elegibilidade ao *Medicare* reduz ou aumenta as disparidades no uso de diferentes tipos de serviços.

Medicare: Definição do modelo

Modelo:

$$y_{ija} = X_{ija}\alpha + f_j(\alpha;\beta) + \sum_k C_{ija}^k \gamma^k + u_{ija}$$

- y_{ija}: uso de cuidados de saúde para o indivíduo i no grupo socioeconômico j na idade a;
- X_{ija}: conjunto de covariáveis (por exemplo, gênero e região);
- $f_j(\alpha;\beta)$: função suavizada representando o perfil de idade do resultado y para o grupo j;
- C_{ija}^k: características da cobertura de seguro mantida pelo indivíduo;
- u_{ija}: componente de erro não observado.

Medicare: Definição do modelo

Problema na estimação (Cobertura do seguro é endógena): a elegibilidade ao programa está associada a uma redução das diferenças de cobertura entre os grupos demográficos, mas há um aumento nessas diferenças quando olhamos para coberturas com mais benefícios.

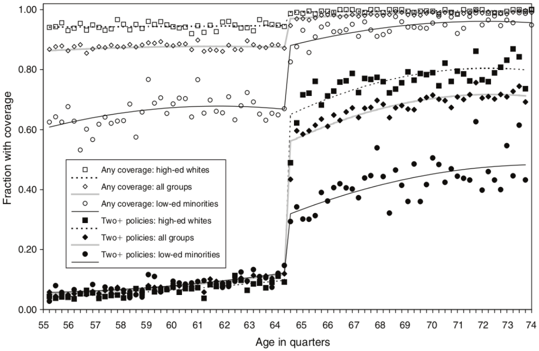


Figura: Cobertura por qualquer seguro e por duas ou mais apólices, por idade e grupo demográfico.

Fonte: Card et al. [2008]

Medicare: Definição do modelo

Solução: Definir um modelo de probabilidade para as variáveis indicadoras C^1_{ija} (qualquer cobertura) e C^2_{ija} (plano com maior cobertura e com mais benefícios):

$$C^1_{ija} = X_{ija}\beta^1_j + g^1_j(a) + D_a\pi^1_j + \nu^1_{ija},$$

$$C^2_{ija} = X_{ija}\beta^2_j + g^2_j(a) + D_a\pi^2_j + \nu^2_{ija},$$

onde β^1_j e β^2_j são coeficientes dos grupos socioeconômicos, $g^1_j(a)$ e $g^2_j(a)$ são perfis de idade destes grupos, e D_a uma indicadora para ter 65 anos ou mais. Supondo que os perfis sejam contínuos aos 65 anos, qualquer descontinuidade em y pode ser atribuída a descontinuidades no seguro.

Medicare: Resultados

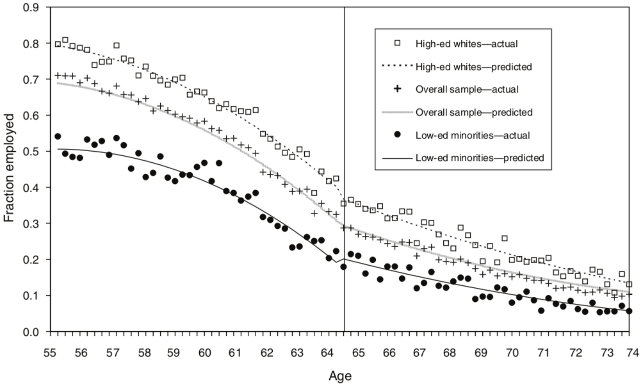
Mudanças na cobertura do *Medicare* aos 65 anos:

- A cobertura do aumenta em 60% aos 65 anos;
- A cobertura antes dos 65 anos é maior para pessoas com escolaridade abaixo da média, e esses grupos experimentam ganhos menores aos 65 anos;
- Ainda existe alguma diferença na cobertura após os 65 anos, mas a diferença de 28% entre brancos com maior escolaridade e minorias com menor escolaridade diminui para cerca de 10%;
- Antes dos 65 anos, as taxas de cobertura privada variam entre 33% para minorias com menor escolaridade e 86% para brancos com melhor escolaridade (essas diferenças dificilmente são afetadas - a maioria dessas pessoas fazem a transição para uma combinação do *Medicare* e cobertura suplementar).

Medicare: Resultados

Outras mudanças aos 65 anos (aposentadoria):

- A continuidade exige que todos os outros fatores que possam afetar o resultado tenham mudanças suaves aos 65 anos.
- Todos os perfis possuem comportamento suave aos 65 anos em relação à empregabilidade.



Fonte: Card et al. [2008]

Medicare: Resultados

	1997–2003 NHIS				1992–2003 NHIS			
	Delayed care last year		Did not get care last year		Saw doctor last year		Hospital stay last year	
	Age 63–64 (1)	RD at 65 (2)	Age 63–64 (3)	RD at 65 (4)	Age 63–64 (5)	RD at 65 (6)	Age 63–64 (7)	RD at 65 (8)
Overall sample	7.2	−1.8	4.9	−1.3	84.8	1.3	11.8	1.2
		(0.4)		(0.3)		(0.7)		(0.4)
Classified by ethnicity and education:								
White non-Hispanic:								
High school dropout	11.6	−1.5	7.9	−0.2	81.7	3.1	14.4	1.6
		(1.1)		(1.0)		(1.3)		(1.3)
High school graduate	7.1	0.3	5.5	−1.3	85.1	−0.4	12.0	0.3
		(2.8)		(2.8)		(1.5)		(0.7)
At least some college	6.0	−1.5	3.7	−1.4	87.6	0.0	9.8	2.1
		(0.4)		(0.3)		(1.3)		(0.7)
Minority:								
High school dropout	13.6	−5.3	11.7	−4.2	80.2	5.0	14.5	0.0
		(1.0)		(0.9)		(2.2)		(1.4)
High school graduate	4.3	−3.8	1.2	1.5	84.8	1.9	11.4	1.8
		(3.2)		(3.7)		(2.7)		(1.4)
At least some college	5.4	−0.6	4.8	−0.2	85.0	3.7	9.5	0.7
		(1.1)		(0.8)		(3.9)		(2.0)
Classified by ethnicity only:								
White non-Hispanic	6.9	−1.6	4.4	−1.2	85.3	0.6	11.6	1.3
		(0.4)		(0.3)		(0.8)		(0.5)
Black non-Hispanic (all)	7.3	−1.9	6.4	−0.3	84.2	3.6	14.4	0.5
		(1.1)		(1.1)		(1.9)		(1.1)
Hispanic (all)	11.1	−4.9	9.3	−3.8	79.4	8.2	11.8	1.0
		(0.8)		(0.7)		(0.8)		(1.6)

Tabela: Medidas de acesso aos cuidados de saúde pouco antes dos 65 anos e descontinuidade estimadas.

Fonte: Card et al. [2008]

Medicare: Resultados

Mudanças no acesso e utilização de cuidados médicos:

- Cerca de 7% das pessoas relataram atrasar os cuidados e 5% relataram não receber cuidados, com taxas mais elevadas para as minorias com menor escolaridade e hispânicos. As estimativas implicam redução aos 65 anos em ambas as medidas;
- Os grupos com menor escolaridade e minoritários têm menor probabilidade de ter uma consulta de rotina, mas são mais propensos a ter passado por um período hospitalar;
- As estimativas sugerem que o limiar dos 65 anos está associado a um aumento nas consultas médicas de rotina, com ganhos maiores para os grupos com taxas mais baixas antes dos 65 anos;
- No geral, há um aumento grande nas taxas de hospitalização aos 65 anos (da ordem dos 10%), mas os ganhos são maiores para os brancos com melhor escolaridade do que para outros grupos.

Pacotes - Linguagem R

- rddtools
- rdd
- rdrobust
- rddensity

Referências (1)

Joshua D Angrist and Jörn-Steffen Pischke. *Mostly harmless econometrics: An empiricist's companion*. Princeton university press, 2009.

David Card, Carlos Dobkin, and Nicole Maestas. The impact of nearly universal insurance coverage on health care utilization: evidence from medicare. *American Economic Review*, 98 (5):2242–2258, 2008.

Thomas D Cook. "waiting for life to arrive": a history of the regression-discontinuity design in psychology, statistics and economics. *Journal of Econometrics*, 142(2):636–654, 2008.

Scott Cunningham. *Causal inference: The mixtape*. Yale university press, 2021. URL `https://mixtape.scunning.com/06-regression_discontinuity`. Acessado: 18/06/2024.

Michael O Finkelstein, Bruce Levin, and Herbert Robbins. Clinical and prophylactic trials with assured new treatment for those at greater risk: I. a design proposal. *American Journal of Public Health*, 86(5):691–695, 1996.

Arthur S Goldberger. Selection bias in evaluating treatment effects: Some formal illustrations. Manuscrito não publicado, 1972a.

Referências (2)

Arthur S Goldberger. Selection bias in evaluating treatment effects: the case of interaction. Manuscrito não publicado, 1972b.

Guido W Imbens and Thomas Lemieux. Regression discontinuity designs: A guide to practice. *Journal of econometrics*, 142(2):615–635, 2008.

David S Lee and Thomas Lemieux. Regression discontinuity designs in economics. *Journal of economic literature*, 48(2):281–355, 2010.

Justin McCrary. Manipulation of the running variable in the regression discontinuity design: A density test. *Journal of econometrics*, 142(2):698–714, 2008.

Donald B Rubin. Assignment to treatment group on the basis of a covariate. *Journal of educational Statistics*, 2(1):1–26, 1977.

Donald L Thistlethwaite and Donald T Campbell. Regression-discontinuity analysis: An alternative to the ex post facto experiment. *Journal of Educational psychology*, 51(6):309, 1960.

William MK Trochim. *Research design for program evaluation: The regression-discontinuity approach*, volume 6. SAGE Publications, Incorporated, 1984.

Perguntas?