

Fundamentos da Regressão

Raphael Corbi

Universidade de São Paulo

Agosto 2017

Propriedades Mecânicas da Regressão

- ▶ regressão são úteis para estimar efeito tratamento de um experimento
 - ▶ com ou sem variáveis de controle!
- ▶ sem tratamento aleatório, a interpretação causal não está garantida
- ▶ essa é a questão chave do curso!
- ▶ mas antes... rever algumas propriedades da regressão
- ▶ independente da interpretação dos coeficientes

Relações econômicas e Esperança Condicional

- ▶ diferenças no desempenho econômico dos indivíduos é aleatória
- ▶ regressões sistematizam aleatoriedade
 - ▶ relação entre educação e escolaridade
 - ▶ fato: pessoas mais educadas ganham mais!
 - ▶ independente de ser ou não uma relação causal
- ▶ educação estatisticamente prevê salário
- ▶ esse poder de previsão fica claro ao vermos que:
 - ▶ regressão $\approx$ esperança condicional

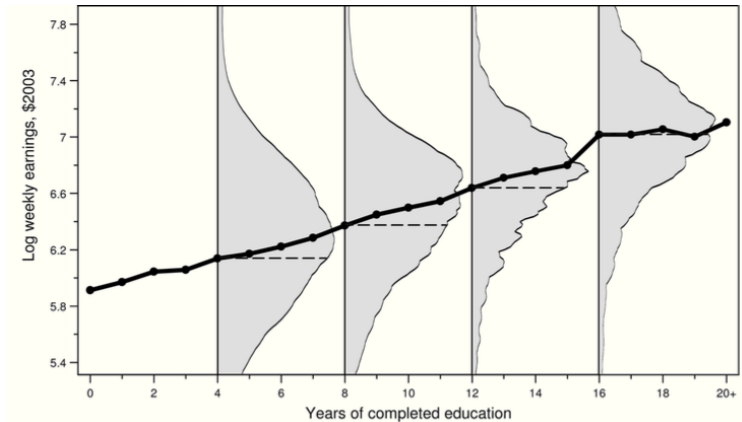
Relações econômicas e Esperança Condicional

$$E[Y_i|X_i = x] = \int tf_y(t|X_i = x)dt$$

$$E[Y_i|X_i = x] = \sum_t tP(Y_i = t|X_i = x)dt$$

- ▶ esperança é um conceito populacional
- ▶ por enquanto, não fazemos a distinção
- ▶ inferimos a EC populacional da EC amostral

Relações econômicas e Esperança Condicional



Lei das Expectativas Iteradas

- ▶ a LEI mostra que uma esperança incondicional pode ser escrita como uma média incondicional da esperança condicional:

$$E[Y_i] = E\{E[Y_i|X_i]\}$$

- ▶ seja (X_i, Y_i) contínuos, $f_{xy}(u, t)$ distribuição conjunta, $f_y(t|X_i = u)$ distribuição condicional de Y_i dado $X_i = u$ e $g_y(t)$ e $g_x(u)$ as densidades marginais

Lei das Expectativas Iteradas - prova

$$\begin{aligned}E[Y_i] &= E\{E[Y_i|X_i]\}\\E\{E[Y_i|X_i]\} &= \int E[Y_i|X_i = u]g_x(u)du\\&= \int \left[\int tf_y(t|X_i = u)dt \right] g_x(u)du\\&= \int t \left[\int f_y(t|X_i = u)g_x(u)du \right] dt\\&= \int t \left[\int f_{xy}(u, t)du \right] dt\\&= \int tg_y(t)dt = E[Y_i]\end{aligned}$$

Propriedade da decomposição da FEC

- ▶ a LEI nos permite separar uma variável aleatória em 2 pedaços, a FEC (explicado por X) e um resíduo ortogonal à qualquer $h(X_i)$:

$$Y_i = E[Y_i|X_i] + \epsilon_i$$

- ▶ (i) ϵ_i é mean-independent de X_i ou seja $E[\epsilon_i|X_i] = 0$,
 - ▶ $E[\epsilon_i|X_i] = E[Y_i - E[Y_i|X_i]|X_i] = E[Y_i|X_i] - E[Y_i|X_i] = 0$
- ▶ (ii) ϵ_i é não-correlacionado com qualquer função de X_i .
 - ▶ $E[h(X_i)\epsilon_i] = E\{h(X_i)E[\epsilon_i|X_i]\}$, e dado (i), $E[\epsilon_i|X_i]=0$

Outras Propriedades da FEC

1) FEC é a melhor previsão de Y_i dado X_i

- ▶ intuição: FEC resolve um problem de mínimo erro quadrado médio

Seja $m(X_i)$ uma função qualquer de X_i . A CEF resolve:

$$E[Y_i|X_i] = \operatorname{argmin} E[(Y_i - m(X_i))^2]$$

Outras Propriedades da FEC

2) Decomposição ANOVA: $Var(Y_i) = Var(E[Y_i|X_i]) + E[Var(Y_i|X_i)]$

- ▶ intuição: $Var(Y_i)$ é decomposto no que pode ser explicado pelas observadas, e pelo resíduo
- ▶ todas estas propriedades valem tanto para população quanto para amostra
- ▶ nada até aqui assumiu linearidade da FEC
- ▶ validade da regressão tampouco requer linearidade da CEF

Regressão Linear e FEC

- ▶ então o que quer dizer: regressão $\approx$ CEF ?
- ▶ seja β um vetor $k \times 1$ definido como: $\beta = \underset{b}{\operatorname{argmin}} E[(Y_i - X_i' b)^2]$
- ▶ pela condição de primeira ordem:

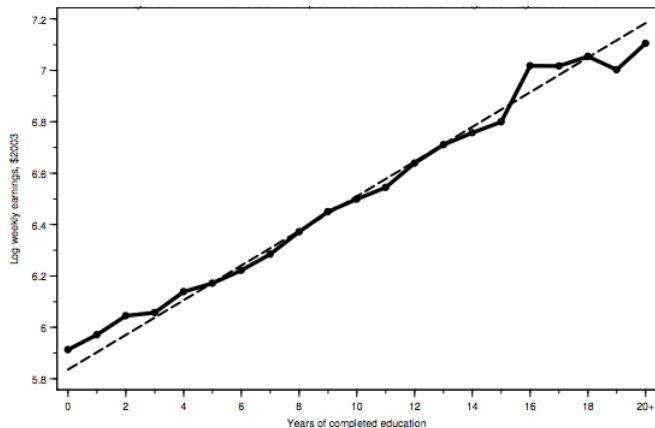
$$E[X_i' \underbrace{(Y_i - X_i' b)}_{\epsilon_i}] = 0$$

- ▶ a solução é $\beta = E[X_i X_i']^{-1} E[X_i' Y_i]$
- ▶ note ϵ_i que não é observado, ele é uma função de β
- ▶ na regressão linear, os resíduos são "forçados" a respeitarem a ortogonalidade em relação a X_i
- ▶ o que acontece se a regressão estiver mal especificada?

Porque usamos a Regressão Linear?

- ▶ Teorema da FEC Linear: Caso FEC linear, regressão=CEF
 - ▶ quando a FEC é linear? (i) $f_{x,y}$ normal, (ii) modelo saturado
- ▶ Teorema: $X_i\beta$ é a melhor previsão linear de $E[Y_i|X_i]$
 - ▶ FEC é a melhor previsão irrestrita, e a regressão é a melhor aproximação linear de $Y_i|X_i$

Porque usamos a Regressão Linear?



Regressão e Causalidade

- ▶ quando um coeficiente de uma regressão captura um efeito causal?
- ▶ relação escolaridade e salário causal se:
 - ▶ podemos mudar escolaridade de maneira perfeitamente controlada,
 - ▶ escolaridade for distribuída aleatoriamente
- ▶ Hipótese da Independência Condicional (HIC)
 - ▶ controlando por observáveis, tratamento é aleatório
 - ▶ seleção-em-observáveis

Hipótese da Independência Condicional

- ▶ Resultado Potencial $Y_i = 0, 1$

$$Y_i = \begin{cases} Y_{1,i} & \text{if } D_i = 1 \\ Y_{0,i} & \text{if } D_i = 0 \end{cases}$$

$$Y_i = Y_{0,i} + (Y_{1,i} - Y_{0,i})D_i$$

- ▶ onde D_i é a decisão de fazer faculdade
- ▶ nunca observamos ambos $Y_{0,i}$ e $Y_{1,i}$

Hipótese da Independência Condicional

$$\underbrace{E[Y_i|D_i = 1] - E[Y_i|D_i = 0]}_{\text{diferença observada em salário médio}} = \underbrace{E[Y_{1,i}|D_i = 1] - E[Y_{0,i}|D_i = 1]}_{\text{efeito tratamento médio nos tratados}} + \underbrace{E[Y_{0,i}|D_i = 1] - E[Y_{0,i}|D_i = 0]}_{\text{viés de seleção}}$$

- ▶ viés de seleção aqui deve ser positivo (superestima)
- ▶ $HIC = \{Y_{0,i}, Y_{1,i}\} \perp\!\!\!\perp D_{0,i} | X_i$
- ▶ $E[Y_{0,i}|X_i, D_i = 1] - E[Y_{0,i}|X_i, D_i = 0] = 0$
- ▶ no caso do Tennessee STAR experiment, qual era o X_i ?

HIC na regressão

- ▶ Seleção nas observáveis

$$Y_i = \alpha + \beta X_i + \eta_i$$

$$\eta_i = Z_i' \gamma + \xi_i$$

$$Y_i = \alpha + \beta X_i + Z_i' \gamma + \xi_i$$

- ▶ mas e se não observarmos Z_i ?

Viés de Variável Omitida

- ▶ regressão correta é $Y_i = \alpha + \beta X_i + Z_i\gamma + \xi_i$
- ▶ como não observamos Z_i , rodamos $Y_i = \alpha + \beta X_i + \xi_i$

$$\begin{aligned}\hat{\beta} &= (X_i'X_i)^{-1}(X_i'Y_i) \\ &= (X_i'X_i)^{-1}X_i'X_i\beta + (X_i'X_i)^{-1}X_i'Z_i\gamma + (X_i'X_i)^{-1}X_i'\xi_i \\ &= \beta + \underbrace{(X_i'X_i)^{-1}X_i'Z_i\gamma}_{\text{viés}}\end{aligned}$$

Exemplo viés variável omitida

Table 3.2.1: Estimates of the returns to education for men in the NLSY

	(1)	(2)	(3)	(4)	(5)
Controls:	None	Age dummies	Col. (2) and additional controls*	Col. (3) and AFQT score	Col. (4), with occupation dummies
	0.132	0.131	0.114	0.087	0.066
	(0.007)	(0.007)	(0.007)	(0.009)	(0.010)

Mas quando a HIC é boa?

- ▶ ex: Black et al. (2003)
- ▶ programa de treinamento obrigatório para desempregados
- ▶ elegibilidade baseada em histórico de emprego e socioeconômico
- ▶ se demanda maior que vagas, loteria
- ▶ regressão dummy treinamento + controles = HIC
- ▶ mas qualquer controle é um bom controle?

Quando/Quais controles devem ser incluídos?

- ▶ todas que foram fixadas ao momento que o regressor foi determinado
- ▶ NÃO incluir se a variável for também dependente!!
- ▶ ex: ensino superior e salário (white/blue collar)
- ▶ devemos controlar por ocupação?
- ▶ e se ensino superior for aleatório para a amostra como um todo?

Qualquer Controle é um Bom Controle?

Seja W_i um indicador de white collar (ocupação):

$$Y_i = C_i Y_{1,i} + (1 - C_i) Y_{0,i}$$
$$W_i = C_i W_{1,i} + (1 - C_i) W_{0,i}$$

onde $C_i = 1$ para universidade (aleatoriamente distribuída)

- ▶ podemos estimar os ATE ao regredir Y_i, W_i em C_i

Qualquer Controle é um Bom Controle?

► modelo de salário e educação condicional à ocupação.

- (i) regressão que controla por W_i
- (ii) restringir a amostra à $W_i = 1^*$

$$\begin{aligned} E[Y_i | W_i = 1, C_i = 1] - E[Y_i | W_i = 1, C_i = 0] &= \\ E[Y_{1i} | W_{1i} = 1, C_i = 1] - E[Y_{0i} | W_{0i} = 1, C_i = 0] &= \\ E[Y_{1i} | W_{1i} = 1] - E[Y_{0i} | W_{0i} = 1] &= \end{aligned}$$

$$\underbrace{E[Y_{1i} - Y_{0i} | W_{1i} = 1]}_{\text{efeito causal nos tratados}} + \underbrace{E[Y_{0i} | W_{1i} = 1] - E[Y_{0i} | W_{0i} = 1]}_{\text{viés de seleção}}$$

Proxies são controles bons?

- ▶ regressão de interesse é $Y_i = \alpha + \beta s_i + a_i \gamma + \epsilon_i$
 - ▶ onde a_i é habilidade (QI) medida antes da faculdade
- ▶ não temos dados sobre a_i , mas temos sobre uma medida de QI a_i^* de uma prova dada ao fim da faculdade
- ▶ em geral, teste de QI reflete inteligência e escolaridade:
 - ▶ $a_i^* = \pi_0 + \pi_1 s_i + \pi_2 a_i$
- ▶ tentando evitar o viés da variável omitida, incluimos a_i^* :
 - ▶ regressão correta é $Y_i = (\alpha - \gamma \frac{\pi_0}{\pi_2}) + (\beta - \gamma \frac{\pi_1}{\pi_2}) s_i + (\frac{\gamma}{\pi_2}) a_i^* + \epsilon_i$

Proxies são controles bons?

- ▶ proxy são bons controles somente se determinados antes do regressor de interesse
 - ▶ em outras palavras, se $\pi_1 = 0$
 - ▶ timing é muito importante aqui (nem sempre óbvio)
- ▶ caso π_1 positivo, sabemos que o coeficiente verdadeiro está entre a estimativa sem controles e a com a proxy mal especificada

Regressão e Matching

- ▶ vimos que se a FEC é linear, a regressão = FEC (ou melhor aprox.)
- ▶ caso a FEC for causal, a regressão será causal?
- ▶ regressão vista como *estimador matching*
- ▶ *ambas técnicas se baseiam na HIC para identificação*
- ▶ *ex: efeito do serviço militar voluntário no salário (Angrist, 1998)*
 - ▶ *HIC: condicional nas observáveis, SM é independente de salário*

Estimador de Matching

- ▶ *matching compara tratamento/controle por grupos de observáveis*
- ▶ *ex: observar simples diferença de médias entre status de veterano:*

$$\underbrace{E[Y_i|D_i = 1] - E[Y_i|D_i = 0]}_{\text{diferença observada em salário médio}} = \underbrace{E[Y_{1,i}|D_i = 1] - E[Y_{0,i}|D_i = 1]}_{\text{efeito tratamento médio nos tratados}} + \underbrace{E[Y_{0,i}|D_i = 1] - E[Y_{0,i}|D_i = 0]}_{\text{viés de seleção}}$$

Estimador de Matching

- ▶ com a HIC = $\{Y_{0,i}, Y_{1,i}\} \perp\!\!\!\perp D_{0,i} | X_i$, o viés de seleção desaparece ao condicionarmos em X_i
- ▶ o efeito do tratamento sobre tratados

$$\begin{aligned}\delta_{TOT} &= E[Y_{1i} - Y_{0i} | D_i = 1] \\ &= E\{E[Y_{1i} - Y_{0i} | X_i, D_i = 1] | D_i = 1\} \\ &= E\{E[Y_{1i} | X_i, D_i = 1] - E[Y_{0i} | X_i, D_i = 1] | D_i = 1\} \\ \delta_X &= E[Y_i | X_i, D_i = 1] - E[Y_i | X_i, D_i = 0]\end{aligned}$$

Estimador de Matching

- ▶ o análogo amostral é:
- ▶ o efeito do tratamento sobre (i) tratados (soldados) ou (ii) médio total (alistados).

$$E[Y_{1i} - Y_{0i} | D_i = 1] = \sum_x \delta_x P(X_i = x | D_i = 1)$$

$$E[Y_{1i} - Y_{0i}] = \sum_x \delta_x P(X_i = x)$$

- ▶ a contraparte de regressão é:

$$Y_i = \sum_x d_{ix} \alpha_x + \delta_R D_i + \epsilon_i$$

Estimador de Matching vs Regressão

Uncontrolled, matching, and regression estimates of the effects of voluntary military

Race	Average earnings in 1988-1991 (1)	Differences in means by veteran status (2)	Matching estimates (3)	Regression estimates (4)	Regression minus matching (5)
Whites	14537	1233.4 (60.3)	-197.2 (70.5)	-88.8 (62.5)	108.4 (28.5)
Non-whites	11664	2449.1 (47.4)	839.7 (62.7)	1074.4 (50.7)	234.7 (32.5)

► São parecidos, mas diferentes. Porque?

Estimador de Matching vs Regressão

- ▶ regressão é um tipo de estimador de matching
- ▶ qual a intuição?
- ▶ mas qual a diferença? pesos...
 - ▶ matching: distribuição dos controles condicional para os tratados
 - ▶ regressão: média ponderada pela variância (mais peso para grupos com mais variância na probabilidade de tratamento)
- ▶ os estimadores serão iguais somente se tratamento for independente dos controles

Propensity Score

- ▶ na regressão, vimos que omissão de controles independentes do regressor de interesse não causa viés em seu coeficiente
- ▶ **Propensity Score** estende esta idéia para matching
- ▶ **Teorema:** caso resultado potencial seja independente de D_i condicional a X_i , então resultado potencial é independente do tratamento condicionado ao propensity score
$$P(X_i) = E[D_i|X_i] = Prob[D_i = 1|X_i]$$

Propensity Score

- ▶ como na regressão, somente devemos condicionar nos regressores correlacionados com a probabilidade do tratamento
- ▶ além disso, controlar pelo propensity score é suficiente!!!
 - ▶ Passo 1: estimar $P(X_i)$ parametricamente
 - ▶ Passo 2: estimador de matching estratificando $P(X_i)$
- ▶ mais formalmente:

$$E[Y_{1i} - Y_{0i} | D_i = 1]$$

$$= E\{E[Y_i | p(X_i), D_i = 1] - E[Y_i | p(X_i), D_i = 0] | D_i = 1\}$$

Regressão vs Propensity Score

- ▶ qual técnica é a melhor?
 - ▶ PS muda o foco da estimação de $E[Y_i|X_i, D_i]$ para $P(X_i) = E[D_i|X_i]$
 - ▶ porém muitas questões ainda abertas: como estimar $P(X_i)$? Inferência?
- ▶ grande debate teórico Lalonde (1986) e Dehejia e Wahba (1999)
- ▶ dados do National Supported Work (experiência trabalho para trabalhadores sem perspectivas)
- ▶ temos avaliação aleatorizada, então podemos comparar técnicas
- ▶ Lalonde (1986) usa controles observados do PSID/CPS (PME/POF)
- ▶ DW revisitam dados usando matching

Regressão vs Propensity Score

Table 3.3.2: Covariate means in the NSW and observational control samples

Variable	NSW		Full Samples		P-score Screened Samples	
	Treated (1)	Control (2)	CPS-1 (3)	CPS-3 (4)	CPS-1 (5)	CPS-3 (6)
Age	25.82	25.05	33.23	28.03	25.63	25.97
Years of schooling	10.35	10.09	12.03	10.24	10.49	10.42
Black	0.84	0.83	0.07	0.20	0.96	0.52
Hispanic	0.06	0.11	0.07	0.14	0.03	0.20
Dropout	0.71	0.83	0.30	0.60	0.60	0.63
Married	0.19	0.15	0.71	0.51	0.26	0.29
1974 earnings	2,096	2,107	14,017	5,619	2,821	2,969
1975 earnings	1,532	1,267	13,651	2,466	1,950	1,859
Number of Obs.	185	260	15,992	429	352	157

Regressão vs Propensity Score

Table 3.3.3: Regression estimates of NSW training effects using alternate controls

Specification	Full Samples			P-Score Screened Samples	
	NSW	CPS-1	CPS-3	CPS-1	CPS-3
	(1)	(2)	(3)	(4)	(5)
Raw Difference	1,794 (633)	-8,498 (712)	-635 (657)		
Demographic controls	1,670 (639)	-3,437 (710)	771 (837)	-3,361 (811) [139/497]	890 (884) [154/154]
1975 Earnings	1,750 (632)	-78 (537)	-91 (641)	no obs. [0/0]	166 (644) [183/427]
Demographics, 1975 Earnings	1,636 (638)	623 (558)	1,010 (822)	1,201 (722) [149/357]	1,050 (861) [157/162]
Demographics, 1974 and 1975 Earnings	1,676 (639)	794 (548)	1,369 (809)	1,362 (708) [151/352]	649 (853) [147/157]

Porque chamamos Regressão de Regressão/?

