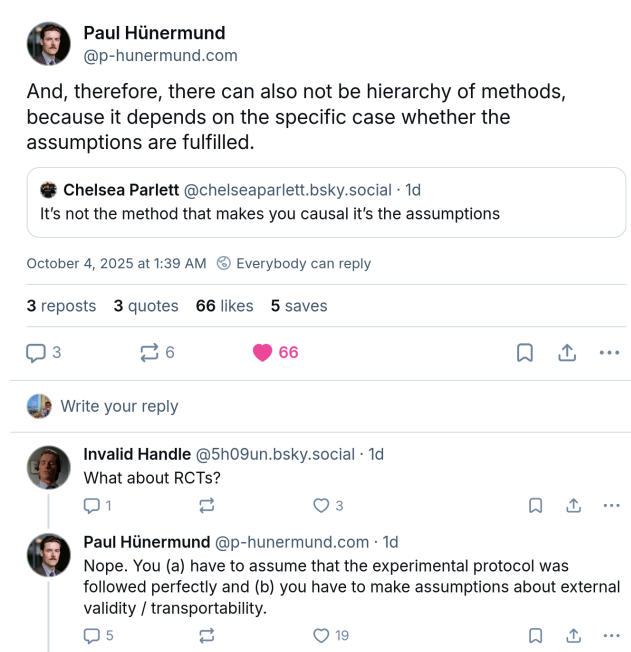


Inferência Causal Baseada em Regressões

IBM0288 - 2026.2

Prof. Raphael Gouvea

1 Para reflexão



2 Tipos de Dados

Em econometria, os dados provêm de duas fontes principais:

- **Dados Experimentais:** Provenientes de experimentos aleatórios controlados (seção anterior). Menos comuns em economia por questões éticas e financeiras.
- **Dados Observacionais:** Obtidos pela observação do comportamento no mundo real (censos, pesquisas, registros administrativos). O desafio é isolar o efeito causal de outros fatores.

3 Estrutura de Dados

Em econometria, os dados são organizados em três estruturas fundamentais:

- + **Corte Transversal (Cross-sectional):** Observações de diferentes entidades (países, estados, indivíduos) em um único período de tempo.

```
# Construindo um data.frame de Corte Transversal
dados_corte_transversal <- data.frame(
  entidade = c("Distrito A", "Distrito B", "Distrito C", "Distrito D",
"Distrito E"),
  nota_teste = c(650, 620, 680, 590, 710)
)

head(dados_corte_transversal)
```

	entidade	nota_teste
1	Distrito A	650
2	Distrito B	620
3	Distrito C	680
4	Distrito D	590
5	Distrito E	710

2. **Séries Temporais:** Observações de uma única entidade ao longo de vários períodos de tempo.

```
# Construindo um data.frame de Séries Temporais
dados_serie_temporal <- data.frame(
  ano = 2010:2015,
  indice_macro = c(102.48, 101.79, 105.03, 112.64, 111.47, 109.13)
)

head(dados_serie_temporal)
```

	ano	indice_macro
1	2010	102.48
2	2011	101.79
3	2012	105.03
4	2013	112.64
5	2014	111.47
6	2015	109.13

Aviso

Séries temporais são conteúdo de Econometria II e não serão tratadas neste curso.

3. **Dados em Painel ou Longitudinais:** Múltiplas entidades, onde cada uma é observada em dois ou mais períodos de tempo.

```
dados_painel <- data.frame(
  pais = rep(c("Argentina", "Brasil"), each = 4),
  ano = rep(2010:2013, 2),
  indice_macro = c(82.25, 81.70, 85.12, 87.50, 102.48, 101.79, 105.03, 112.64)
)

head(dados_painel, n = 8)
```

	pais	ano	indice_macro
1	Argentina	2010	82.25
2	Argentina	2011	81.70
3	Argentina	2012	85.12
4	Argentina	2013	87.50
5	Brasil	2010	102.48
6	Brasil	2011	101.79
7	Brasil	2012	105.03
8	Brasil	2013	112.64

4 Regressão

Chapter 2

Regression



KWAI CHANG CAINE: A worker is known by his tools. A shovel for a man who digs. An ax for a woodsman. The econometrician runs regressions.

Kung Fu, Season 1, Episode 8

5 Educação privada de elite

Vale a pena pagar uma universidade privada de elite? Por quê?

Existe prêmio (diferencial salarial futuro) por estudar em universidade privada de elite?

Como separar efeito causal de viés de seleção?

6 Trade-offs

Custos

- Universidade privada de elite: ~US\$ 29 mil/ano
- Universidade pública estadual: ~US\$ 9 mil/ano
- Diferença: ~US\$ 20 mil/ano
- Investimento total: > US\$ 100 mil

Benefícios

- Turmas menores

- Professores renomados
- Estrutura moderna
- Colegas mais seletos
- *Networking*

7 Questão central

- Vale a pena pagar a diferença de custo?

- Comparações simples (Harvard vs. U-Mass ou IBMEC vs. UnB) são **viesadas**
 - Por quê?
- Problema de seleção: diferenças de notas, renda e motivação entre os admitidos em cada tipo de universidade

Para refletir

- Nancy Qian (profa. de Northwestern): passou em Harvard → escolheu UT por bolsa
- Amanda Pallais (profa. de Harvard): passou em várias elites → escolheu UVA
- Ambas tiveram carreiras de sucesso

8 O que fazer?

Problema

- Diferenças de renda refletem:
 - Tanto a qualidade da instituição
 - Quanto as características dos alunos

Desafio

- Harvard não sorteia vagas
- Não há experimento aleatório que possamos usar

9 Possível solução

- Explorar **quase-aleatoriedade**:
 - Bolsas de estudo
 - Circunstâncias pessoais
 - Timing de decisões
- Resultado: grupos de tratamento e controle obtidos por **experimento natural**

Dica

É muito provável que exista prêmio de estudar em uma escola de elite, mas separar **efeito causal** de **viés de seleção** exige método rigoroso.

10 Notas e Tamanho das Turmas

Reduzir o tamanho das turmas melhora o desempenho escolar dos alunos?

Pergunta de política pública

- **Dilema:** Contratar mais professores para reduzir o número de alunos por sala custa dinheiro dos contribuintes. Esse investimento compensa?

Questão de Inferência Causal: necessidade de isolar e conhecer o *efeito causal* de diminuir a turma em 1 aluno sobre as notas nos testes.

Pergunta da família

- **Dilema:** Escolher o distrito escolar do filho que levará ao melhor desempenho acadêmico possível usando o tamanho da turma como indicador.

Questão de Previsão: não importa o mecanismo causal, a necessidade é prever o resultado a partir de uma variável indicadora

11 Inferência causal e regressões

Estimação:

- Como devemos traçar uma linha através dos dados para estimar o efeito tratamento?
- Quando o parâmetro estimado **identifica** o efeito causal?
- Quais são as vantagens e desvantagens do MQO?

Inferência:

- Como testar se o efeito tratamento é igual a zero?
- Como construir um intervalo de confiança para o efeito tratamento?

! Importante

Inferência causal e previsão impõem requisitos diferentes aos dados, ainda que ambas utilizem regressão.

12 O modelo de regressão linear

$$Y_i = \beta_0 + \beta_1 X_i + u_i$$

- Y : variável dependente
- X : variável independente (regressor)
- $\beta_0 + \beta_1 X_i$: função de regressão populacional
- β_0 : intercepto populacional
- β_1 : inclinação populacional
- u_i : termo de erro populacional

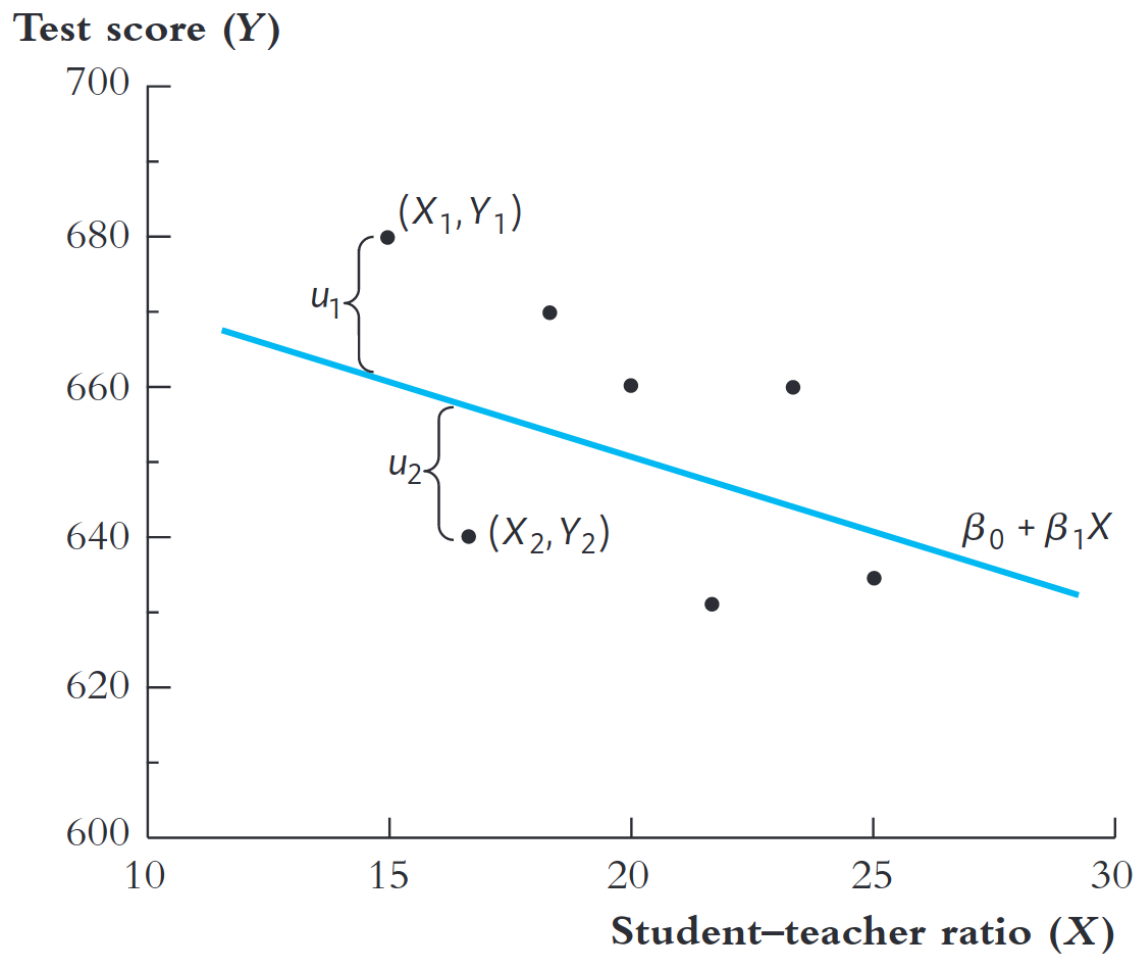
13 Exemplo

Modelo:

$$\text{TestScore}_i = \beta_0 + \beta_1 \text{STR}_i + u_i$$

Média condicional:

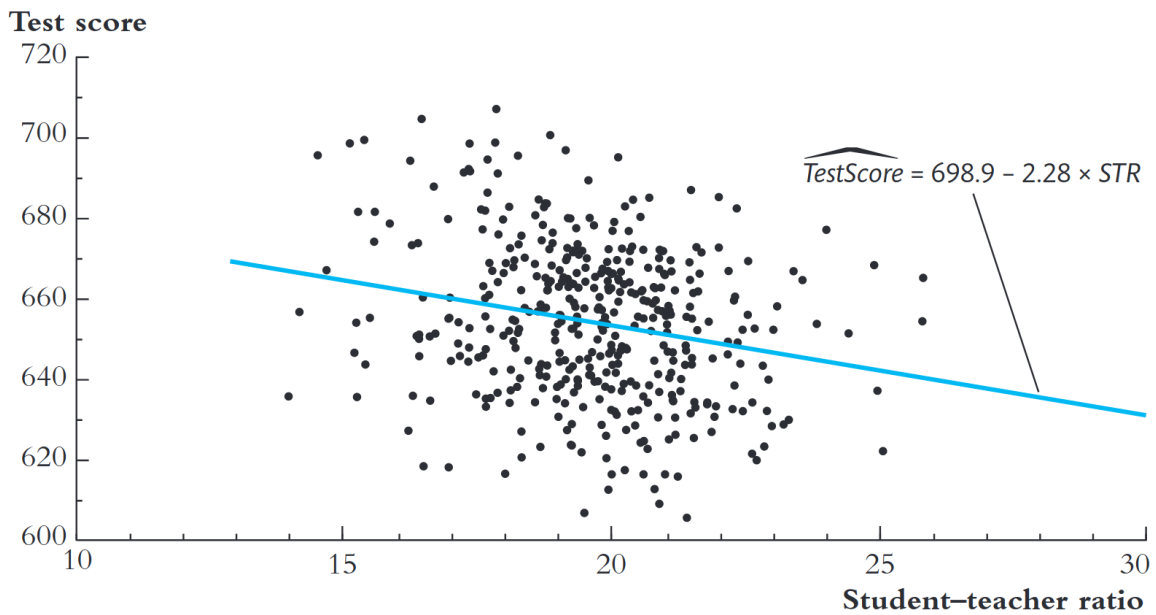
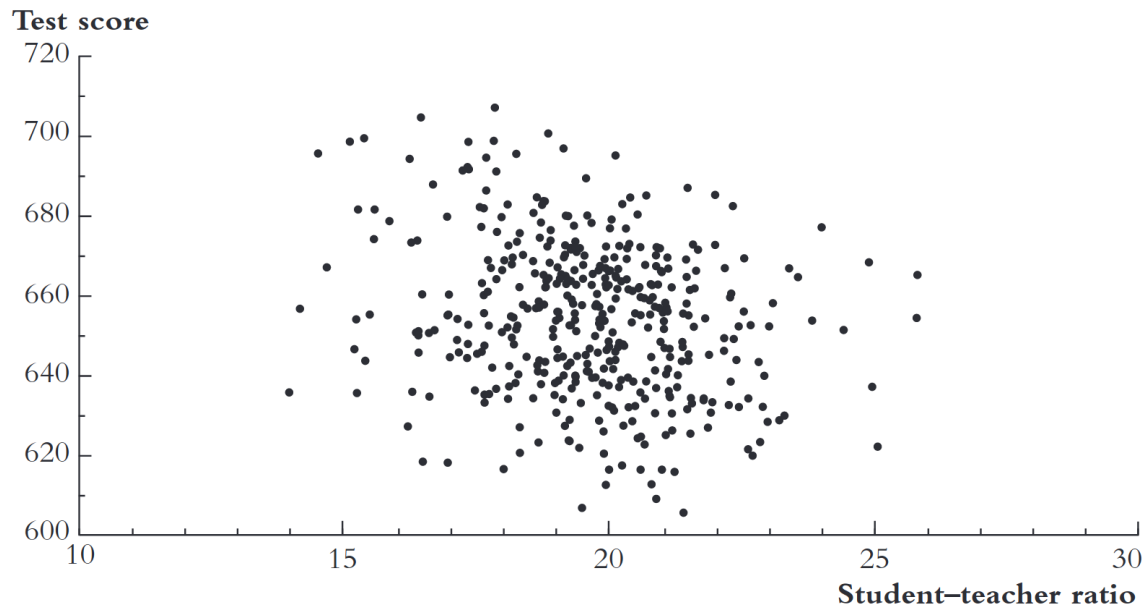
$$E(\text{TestScore} | \text{STR}) = \beta_0 + \beta_1 \text{STR}$$



14 Estimação do modelo de regressão linear

- Podemos estimar β_0 e β_1 a partir de uma amostra.
- Escolher β_0, β_1 para **ajustar melhor** aos dados.

...



15 MQO: um problema de minimização

Melhor ajuste = minimizar erros de previsão ao quadrado.

$$(\hat{\beta}_0, \hat{\beta}_1) = \arg \min_{b_0, b_1} \sum_{i=1}^n (Y_i - [b_0 + b_1 X_i])^2$$

A solução do problema acima define os estimadores de **Mínimos Quadrados Ordinários (MQO)**.

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^n (X_i - \bar{X})^2} = \frac{Cov(X, Y)}{Var(X)}$$

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}$$

16 Os estimadores de MQO

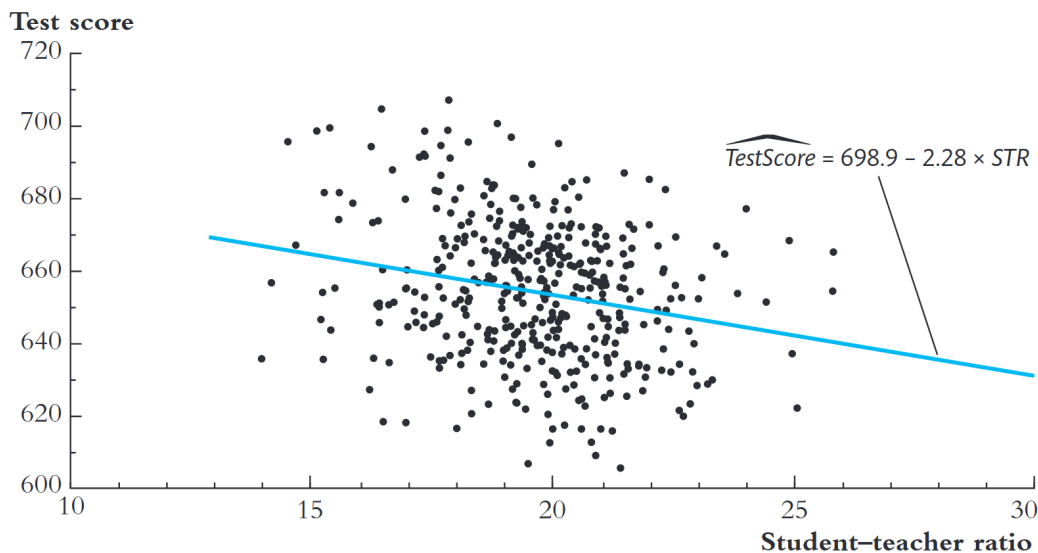
- Modelo de regressão linear: $\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i + \hat{u}_i$

- Os coeficientes $\hat{\beta}_0$ e $\hat{\beta}_1$ são **estimativas pontuais** dos parâmetros populacionais β_0 e β_1 a partir dos dados da amostra
- $\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i$: valor previsto de Y_i com base em X_i .
- $\hat{u}_i$: resíduo da regressão (estimador do erro u_i).

Dica

O erro de regressão captura **todos** os fatores omitidos no modelo. Em geral, esses fatores omitidos são outros fatores que influenciam Y , além da variável X .

17 Estimativas do exemplo



18 Interpretação

- Distritos com um aluno a mais por professor, em média, têm pontuações em testes 2,28 pontos menores:

$$\hat{\beta}_1 = \frac{\Delta E(\text{TestScore} \mid \text{STR})}{\Delta \text{STR}} = -2,28$$

- Note que como $Y = \beta_0 + \beta_1 X$, se X muda por uma quantidade ΔX , então Y muda por: $\Delta Y = \beta_1 \Delta X$

Dica

β_0 é o valor esperado de Y quando $X = 0$. Em muitos modelos econômicos, $X = 0$ não faz sentido prático, então β_0 serve apenas para posicionar a reta de regressão.

19 Ajuste do modelo: R^2

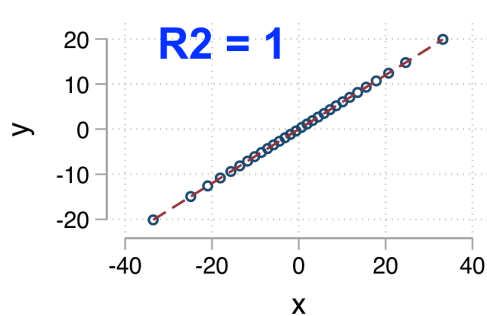
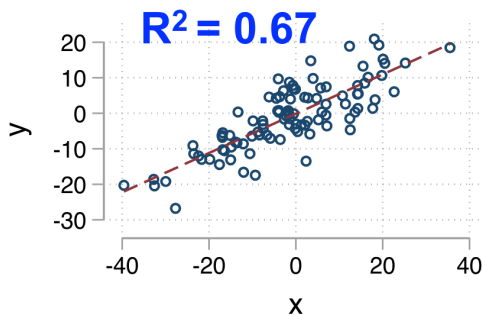
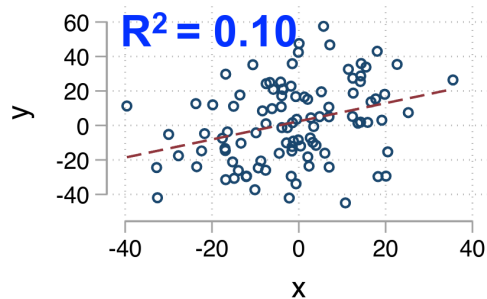
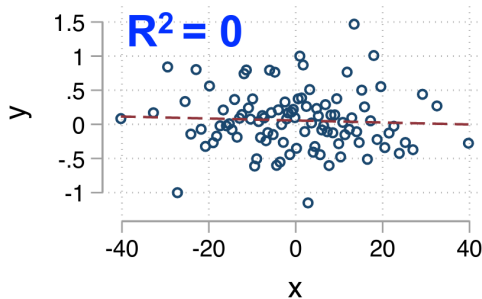
- $Y_i = \hat{Y}_i + \hat{u}_i = \text{Predição de MQO} + \text{Resíduo de MQO}$
- $\text{Var}(Y_i) = \text{Var}(\hat{Y}_i) + \text{Var}(\hat{u}_i)$
- $R^2 = \frac{\text{Var}(\hat{Y}_i)}{\text{Var}(Y_i) + \text{Var}(\hat{u}_i)} = \frac{\text{Var}(\hat{Y}_i)}{\text{Var}(Y_i)} = \frac{\sum_{i=1}^n (\hat{Y}_i - \bar{Y})^2}{\sum_{i=1}^n (Y_i - \bar{Y})^2} = \frac{ESS}{TSS}$

...

ESS: Soma dos quadrados explicados

TSS: Soma dos quadrados totais

20 Visualização de R^2 distintos



! Importante

Não iremos focar nas medidas de ajuste neste curso. Essas medidas têm deixado de ser enfatizadas. O foco é nas hipóteses de **identificação** de causalidade!

21 Ajuste do modelo: SER

- **SER**: *Standard Error of the Regression*
- Mede a dispersão dos resíduos em torno da reta.

$$SER = \hat{\sigma}_u = \sqrt{\frac{1}{n-2} \sum_{i=1}^n \hat{u}_i^2}$$

22 Considerações sobre ajuste de modelo

- Espere R^2 **baixo** quando muitos fatores não observados afetam Y .
- O **SER** informa a magnitude típica do erro de previsão.

! Importante

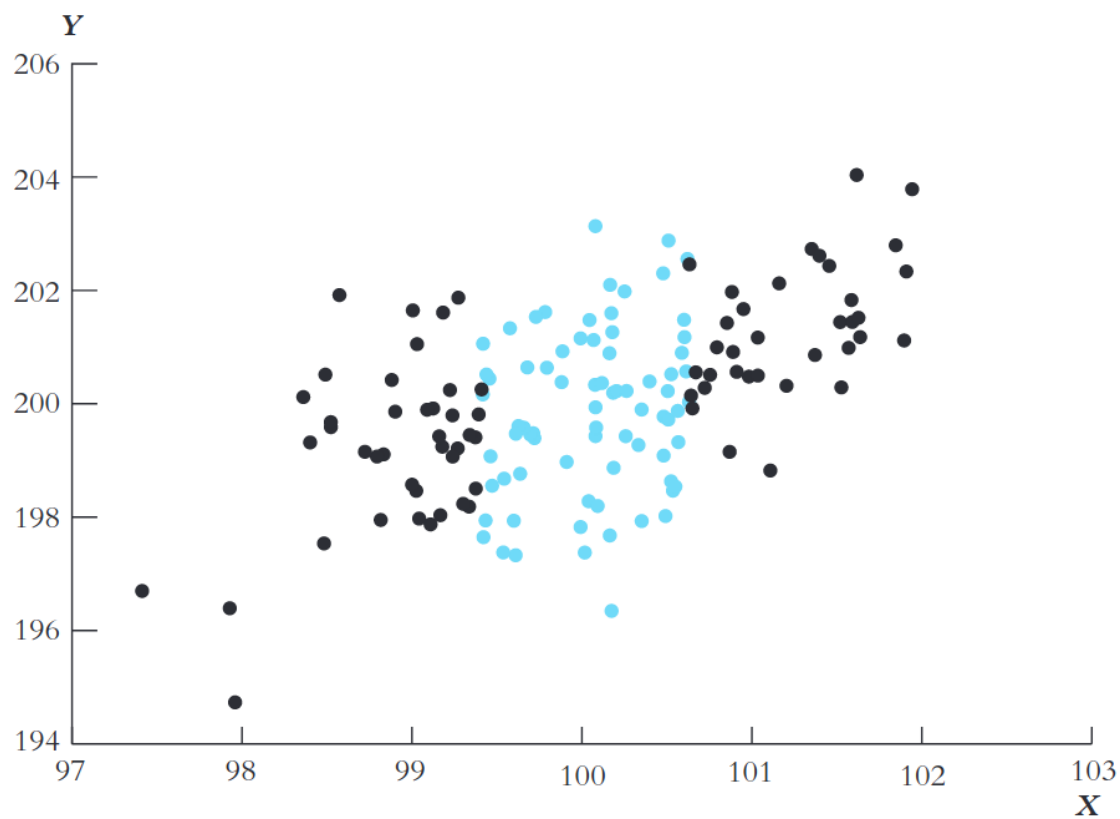
Não iremos focar nas medidas de ajuste neste curso. Essas medidas têm deixado de ser enfatizadas. O foco é nas hipóteses de **identificação** de causalidade!

23 Distribuição amostral dos estimadores

- $\hat{\beta}_0$ e $\hat{\beta}_1$ são variáveis aleatórias (por quê?).
- $E[\hat{\beta}_0] = \beta_0$ e $E[\hat{\beta}_1] = \beta_1$
- $\hat{\beta}_0$ e $\hat{\beta}_1$ são normalmente distribuídos em amostras grandes
- $\hat{\beta}_0 \sim N(\beta_0, \sigma_{\hat{\beta}_0}^2)$ e $\hat{\beta}_1 \sim N(\beta_1, \sigma_{\hat{\beta}_1}^2)$
- O que determina $\sigma_{\hat{\beta}_0}^2$ e $\sigma_{\hat{\beta}_1}^2$?

24 Variância do estimador - intuição

- O número de observações em preto e em azul é igual
- Todas as observações vêm da mesma distribuição conjunta
- Para qual dos dois grupos a reta de regressão é melhor estimada?
- Aumentar a dispersão de X diminui a $var(\hat{\beta}_1)$



25 Homocedasticidade vs Heterocedasticidade

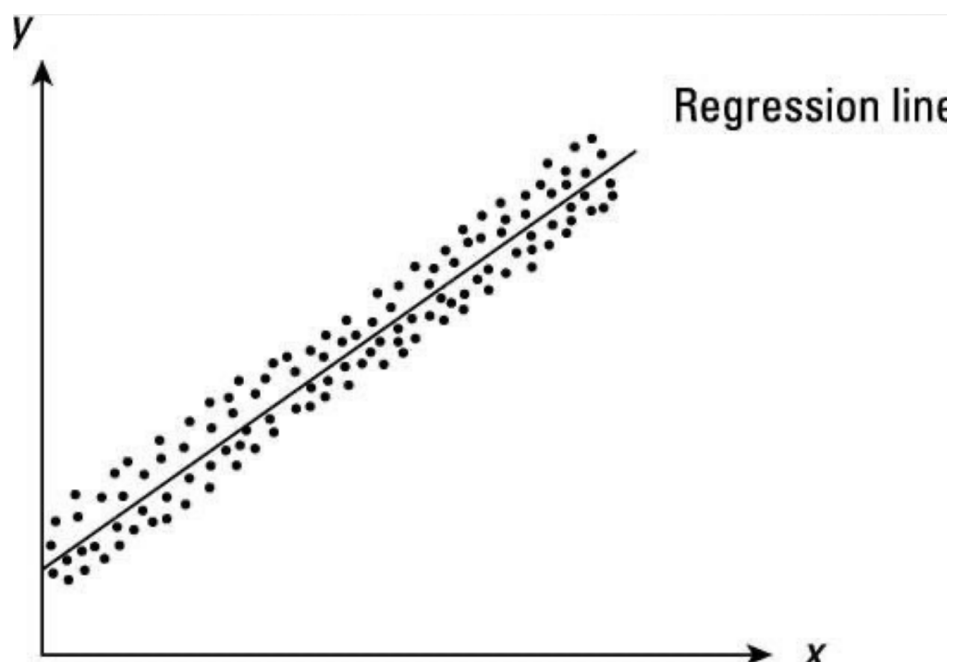
- Conceito relacionado à variância condicional do termo de erro:

...

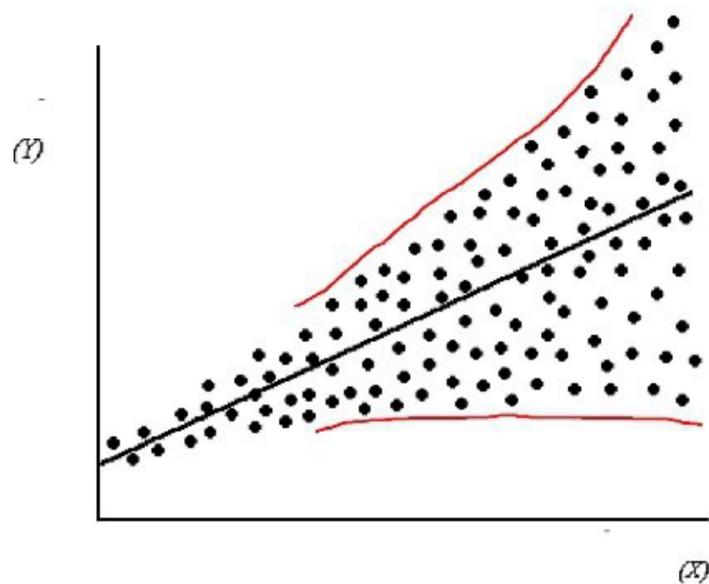
$$\text{var}(u \mid X = x)$$

- Homocedasticidade: $\text{var}(u \mid X = x)$ é constante (**não depende de X**)
- Heterocedasticidade: $\text{var}(u \mid X = x)$ **varia com X**

26 Visualização



Heteroskedasticity



...

Quando devemos usar desvios-padrão sob homocedasticidade (não robustos)?

 Dica

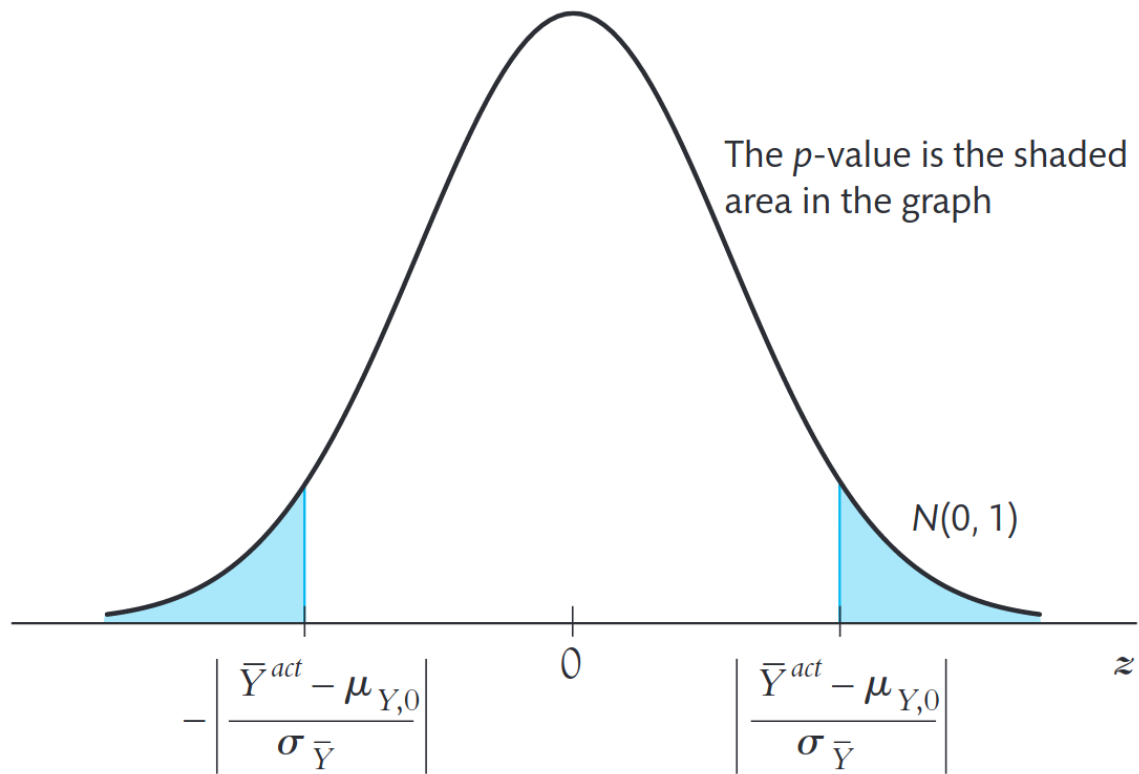
NUNCA! Os desvios-padrão robustos são sempre mais adequados, já que também são válidos sob a hipótese de homocedasticidade.

27 Testes de hipóteses

$$H_0 : \beta_k = \beta_{k,0} \quad \text{vs.} \quad H_1 : \beta_k \neq \beta_{k,0}$$

Passos práticos para testar H_0 :

1. Estime $\hat{\beta}_1$ e $SE(\hat{\beta}_1)$.
2. Calcule a estatística t
3. Calcule o p-valor



28 1. Estime $\hat{\beta}_1$ e $SE(\hat{\beta}_1)$

- $SE(\hat{\beta}_1)$ é o estimador da variância de $\hat{\beta}_1$ ($\sigma_{\hat{\beta}_1}$)

- Erro-padrão robusto (ajuste para heterocedasticidade) - regressão simples:

...

$$SE(\hat{\beta}_1) = \sqrt{\sigma_{\hat{\beta}_1}} = \frac{1}{n} \times \frac{\frac{1}{n-2} \sum_{i=1}^n (X_i - \bar{X})^2 \hat{u}_i^2}{\left[\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2 \right]^2}$$

- Fórmula complicada, mas softwares (R, Stata, etc.) calculam automaticamente.

29 2. Calcule a estatística t

$$t = \frac{\text{estimador} - \text{valor hipotético}}{\text{erro padrão do estimador}}$$

- t tem distribuição normal padrão em grandes amostras
- $t \sim N(0, 1)$

30 3. Calcule o p-valor

Seja $\hat{\beta}_1^{\text{ef}}$ a estimativa efetivamente calculada:

$$\begin{aligned} \text{p-valor} &= \Pr \left(\left| \hat{\beta}_1 - \beta_{1,0} \right| > \left| \hat{\beta}_1^{\text{ef}} - \beta_{1,0} \right| \right) \\ &= \Pr \left(\left| \frac{\hat{\beta}_1 - \beta_{1,0}}{S(\hat{\beta}_1)} \right| > \left| \frac{\hat{\beta}_1^{\text{ef}} - \beta_{1,0}}{S(\hat{\beta}_1)} \right| \right) \\ &= \Pr_{H_0} (|t| > |t^{\text{ef}}|) \\ &= \Pr_{H_0} (|Z| > |t^{\text{ef}}|) \\ &= 2(\phi - |t^{\text{ef}}|) \end{aligned}$$

31 Intervalo de Confiança de 95% para β_k

- Intervalo de Confiança de 95%: intervalo que possui 95% de probabilidade de conter o **verdadeiro** valor de β_k .

- Inclui todos os valores de β_k que não podemos rejeitar ao nível de significância de 5%.

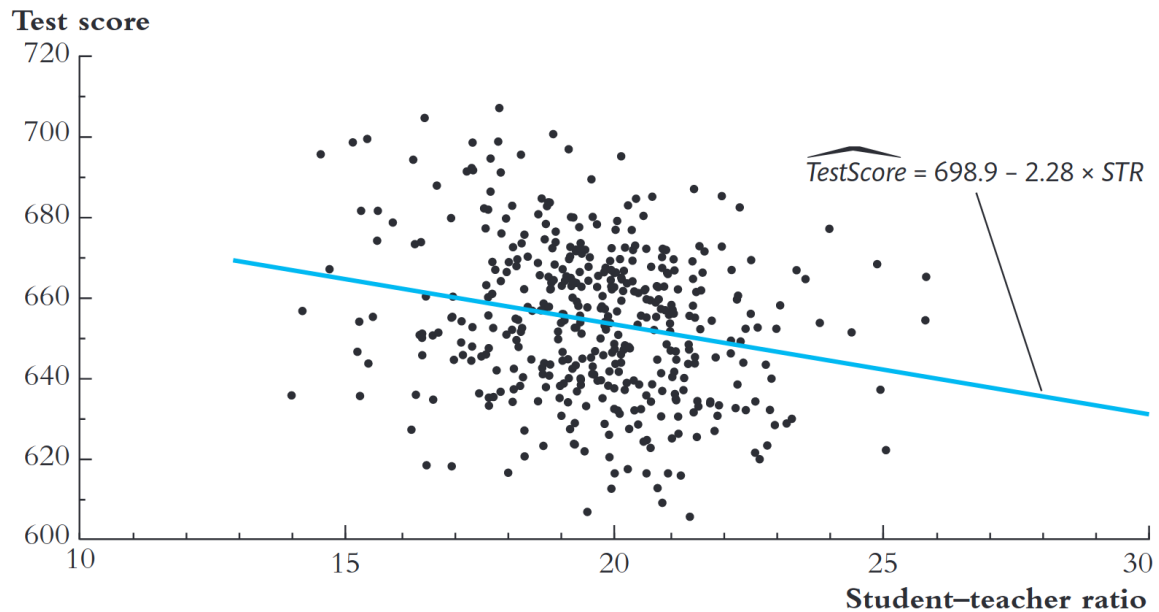
...

$$\hat{\beta}_k - 1,96 \times SE(\hat{\beta}_k) \leq \beta_k \leq \hat{\beta}_k + 1,96 \times SE(\hat{\beta}_k)$$

Aviso

Regra simples: sempre que o módulo da estimativa pontual for duas vezes maior que o desvio-padrão, rejeita-se a hipótese nula.

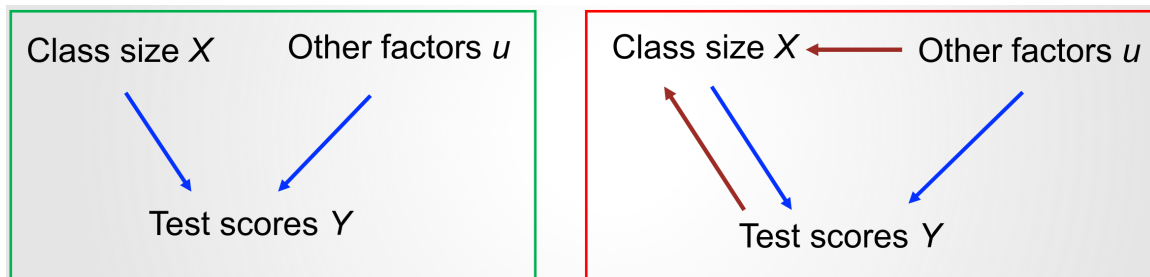
32 Notas e tamanho das turmas



Efeito causal do tamanho da sala?

Ou alguma outra coisa?

33 Relações causais entre notas e tamanho das turmas



Quando uma das setas vermelhas está presente, não é possível garantir que os coeficientes de MQO capturam o efeito causal.

34 Regressões e causalidade

- Quando β_1 pode ser interpretado como efeito causal médio de X sobre Y ?
- X precisa ser independente de outros fatores que afetam Y
 - X tem que ser independente do termo de erro u_i
 - $\text{corr}(X_i, u_i) = 0$
- Isso acontece para dados experimentais!

- Não será sempre verdadeiro para dados observacionais!

35 Três hipóteses de MQO para inferência causal

+ A variável independente X é independente do termo de erro u_i

$$E[u_i | X_i] = 0; \text{corr}(X_i, u_i) = 0$$

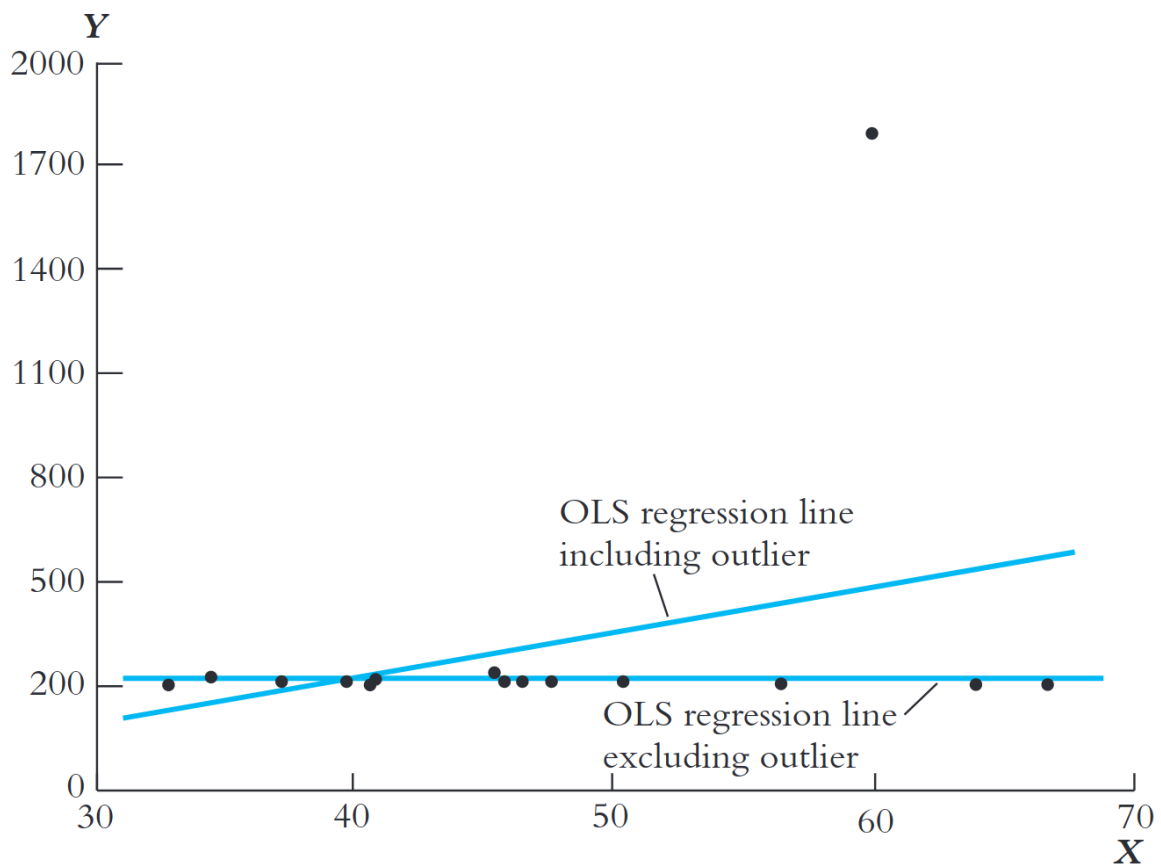
2. (X_i, Y_i) , $i = 1, \dots, n$, são independentes e identicamente distribuídos (i.i.d.).

- Algumas violações de independência podem ser resolvidas com séries temporais ou dados em painel

3. Sem grandes outliers em X e/ou Y .

- A presença de outliers pode distorcer os estimadores de MQO.

36 Sensibilidade a outliers



Pontos extremos em X ou Y podem **distorcer** a reta.