



# DCP098

## Fundamentos para Avaliação Quantitativa de Políticas Públicas

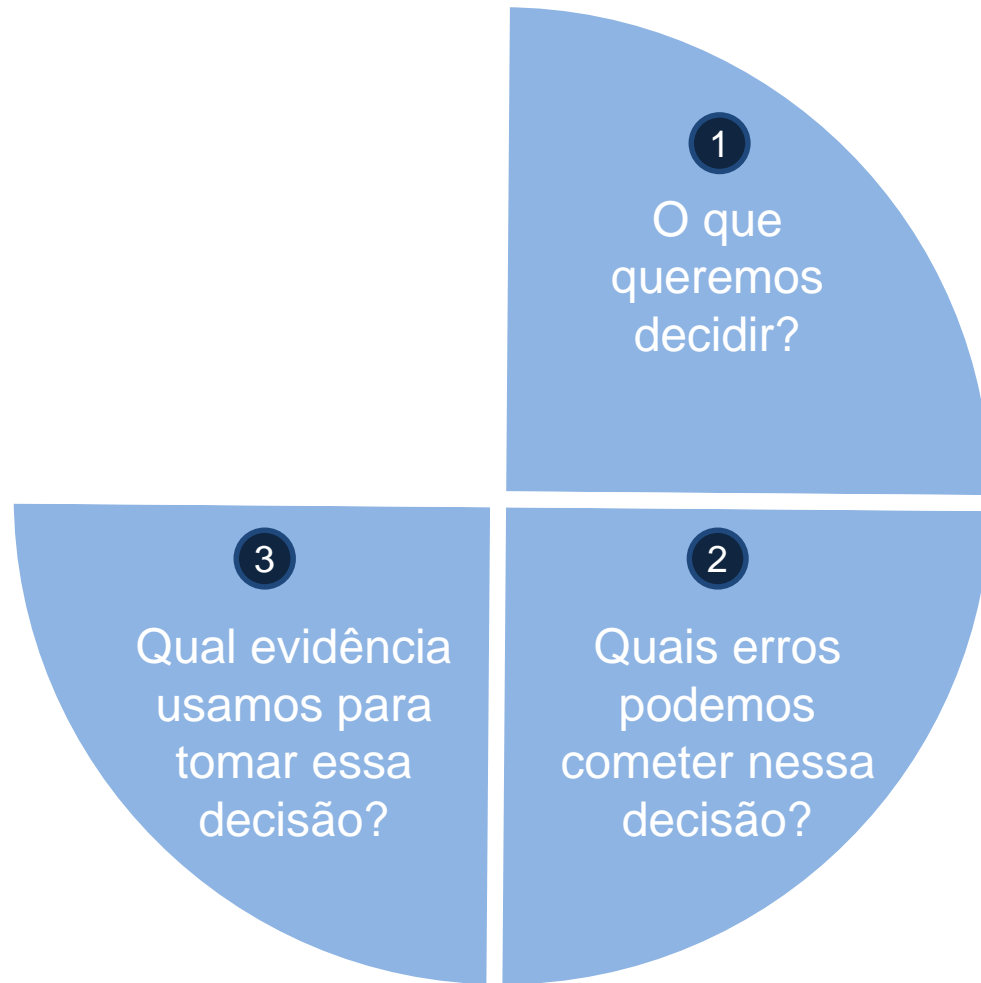
### Teste de hipótese

**Aula 14**  
18 de maio de 2022

Ana Paula Karruz

# Conclusões sobre efeito de X em Y requerem tomada de decisão sobre o significado da evidência apurada

Teste de hipótese é rotina para essa tomada de decisão



# Estamos tentando descobrir se $\beta_j$ seria altamente improvável se $H_0$ for verdadeira

Desafio: Como  $\beta_j$  tem a escala do Y, a escala de  $\beta_j$  poderia ser qualquer uma!

O que nos importa não é o coeficiente estimado  $\beta_j$  em si, mas **quão grande** o coeficiente  $\beta_j$  é em relação a seu erro padrão.

Bailey (2016: 148)

Portanto usamos uma **estatística de teste** que consiste no coeficiente estimado dividido pelo seu erro padrão estimado:

$$\frac{\hat{\beta}_j}{se(\hat{\beta}_j)}$$

Esta razão é chamada de **estatística t**; tem **mesmo sinal de  $\beta_j$** , pois  $se(\beta_j) > 0$

Podemos testar qualquer hipótese do tipo  $H_0: \beta_j = \beta_j^{Nula}$  usando a seguinte razão como estatística de teste:

$$\frac{\hat{\beta}_j - \beta_j^{Nula}}{se(\hat{\beta}_j)}$$

Assim, nossa estatística de teste reflete **quantos erros padrão acima ou abaixo de zero [ou outro valor contido em  $H_0$ ] o coeficiente estimado está.**

Bailey (2016: 148)

# Enquanto $H_0$ define uma média para a distribuição de $\beta_j$ hats, recorremos à amostra para estimar o erro padrão

*Qual é a distribuição de  $\beta_1$ hat **sob a hipótese nula** [ $H_0$ :  $\beta_1 = 0$ ]? Muito simples: É uma variável aleatória normalmente distribuída e **centrada** em zero porque MQO é um estimador não enviesado [dado que suas premissas tenham sido atendidas] e, se o verdadeiro valor de  $\beta_1$  é zero [sob a hipótese nula], então uma distribuição não enviesada de  $\beta_1$ hat estará centrada em zero.*

Bailey (2016: 141)

***Quão larga [espalhada]** é a distribuição de  $\beta_1$ hat sob a hipótese nula? Em contraste com a média da distribuição [media esta] que conhecemos sob a nula, a **largura depende dos dados e do erro padrão** calculado com base nesses dados. Em outras palavras, permitimos aos dados que nos informem sobre o erro padrão de  $\beta_1$ hat [...].*

Bailey (2016: 142)

# Ao ponderarmos $\hat{\beta}_j$ por seu erro padrão, obtemos uma estatística de teste que segue a distribuição t

- $se(\hat{\beta}_j)$  é uma **variável aleatória**, pois depende de uma variável aleatória ( $\hat{\beta}_j$ )
- $se(\hat{\beta}_j)$  corresponde à raiz quadrada de uma distribuição **qui-quadrado** dividida por seus graus de liberdade
- A **razão** entre uma distribuição **normal padronizada** e a raiz quadrada de uma distribuição **qui-quadrado** dividida por seus graus de liberdade (como a distribuição de  $se(\hat{\beta}_j)$ ) corresponde a uma **distribuição t**

$$\hat{\beta}_j / se(\hat{\beta}_j) \sim t$$

*Formalmente, se  $z$  é uma variável aleatória com distribuição normal padrão e  $x$  é uma variável  $\chi^2$  [chi-quadrado] com  $n$  graus de liberdade, então a função abaixo comporta-se como uma distribuição t com  $n$  graus de liberdade:*

$$t(n) = \frac{z}{\sqrt{x/n}}$$

Bailey (2016: 785-6).

- Uma distribuição  $t_{GdeL}$  é **semelhante a uma  $N(\mu, \sigma^2)$** : simétrica, centrada em sua média e em formato de sino; porém, a t possui **caudas mais pesadas**



# DEEP DIVE 1: Por que estatística $t$ apresenta distribuição $t_{n-k-1}$ ?

# DEEP DIVE 1:

## Por que estatística $t$ apresenta distribuição $t_{n-k-1}$ ? (1/3)

- Distribuição de  $\beta_1\text{hat}$ : compreende **todos os possíveis valores** que  $\beta_1\text{hat}$  pode assumir, e a **probabilidade** relativa desses valores
- Essa distribuição, chamada **distribuição amostral de  $\beta_1\text{hat}$** , é uma distribuição “**teórica**”: não nos é visível, pois observamos apenas o  $\beta_1\text{hat}$  gerado por nossa (tipicamente única) amostra
- **Fórmula** do  $\beta_1\text{hat}$  pode ser **reescrita como uma média** (Bailey, 2016: 85, nota 8)
- Como tal,  **$\beta_1\text{hat}$  é uma estatística amostral sujeita ao Teorema Central do Limite** (aka Teorema do Limite Central), segundo o qual:
  - A média amostral de qualquer variável aleatória segue uma distribuição normal, com média = média da distribuição teórica (i.e.,  $\mu = \beta_1$ )
  - Quanto maior o tamanho da amostra, mais próxima de uma normal será a distribuição amostral da média



# DEEP DIVE 1:

## Por que estatística t apresenta distribuição $t_{n-k-1}$ ? (2/3)

- Se o erro estocástico ( $\varepsilon_i$ ) for normalmente distribuído, a variância da distribuição teórica de  $\beta_1$ hat será:

$$\text{var}(\hat{\beta}_1) = \frac{\hat{\sigma}^2}{N \times \text{var}(X)}$$

$$\text{var}(\hat{\beta}_1) = \sigma^2 / \sum (X_i - \bar{X})^2$$

onde  $\sigma^2$  = variância de  $\varepsilon$ , que é desconhecida;

$\sigma^2$  é estimada por  $\hat{\sigma}^2$ , a variância da regressão, que corresponde à soma dos quadrados dos resíduos dividida pelos graus de liberdade do modelo ( $n - j = n - k - 1$ ):

$$\begin{aligned}\hat{\sigma}^2 &= \frac{\sum_{i=1}^N (Y_i - \hat{Y}_i)^2}{N - j} \\ &= \frac{\sum_{i=1}^N \hat{\varepsilon}_i^2}{N - j}\end{aligned}$$

- Quando apuramos a distância entre a evidência amostral ( $\beta_1$ hat) e o cenário da hipótese nula ( $\beta_1^{\text{Nula}}$ ), e ponderamos essa distância pela imprecisão da evidência amostral (i.e., pelo erro padrão de  $\beta_1$ hat, o  $\text{se}(\beta_1$ hat)), estamos calculando a seguinte razão:

$$\beta_1\text{hat} - \beta_1^{\text{Nula}} / \text{se}(\beta_1\text{hat})$$

- Que, na **regressão simples e na regressão múltipla isenta de multicolinearidade**, pode ser reescrita como a razão entre uma distribuição normal padrão e a raiz quadrada de uma distribuição qui-quadrado dividida por seus graus de liberdade (vide próxima página)
- Ou seja, a estatística t se comporta como uma distribuição t com  $n - k - 1$  graus de liberdade



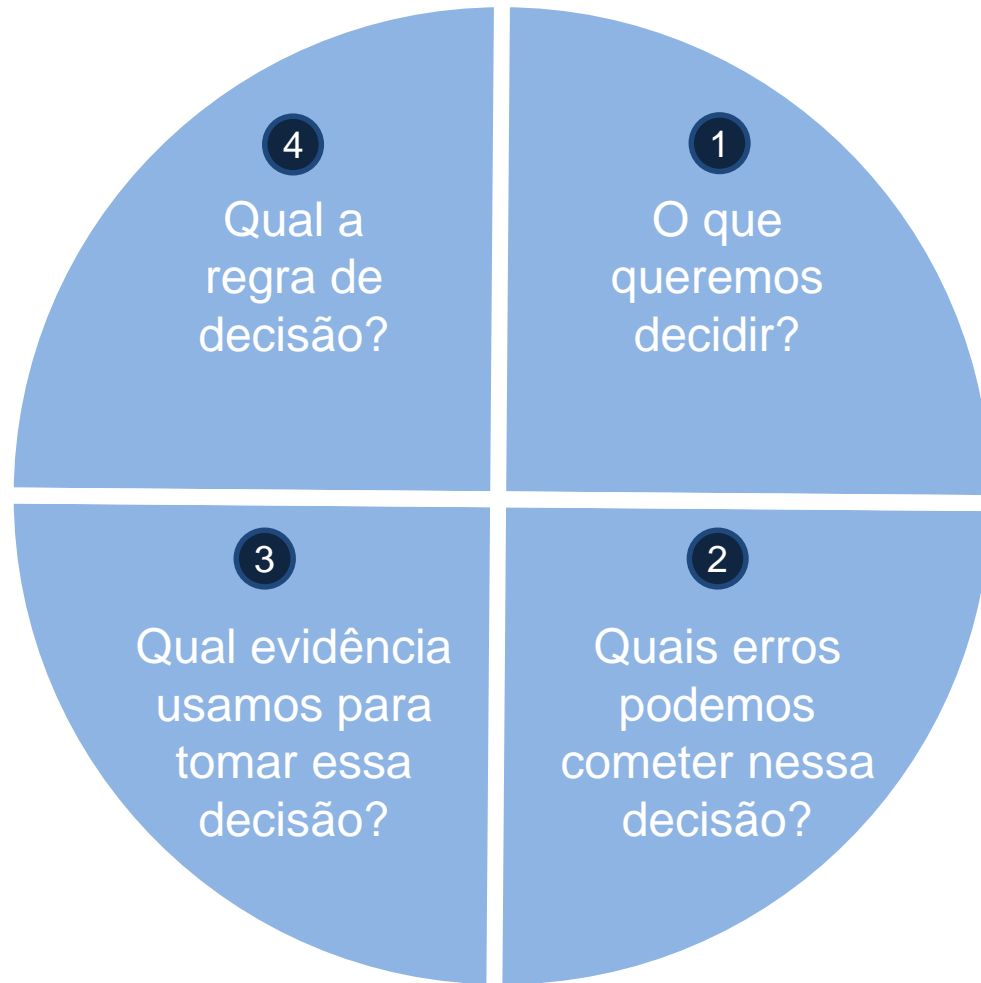




Back to the surface

# Conclusões sobre efeito de X em Y requerem tomada de decisão sobre o significado da evidência apurada

Teste de hipótese é rotina para essa tomada de decisão



# À medida que o tamanho da amostra cresce, a distribuição t tende à normal padrão

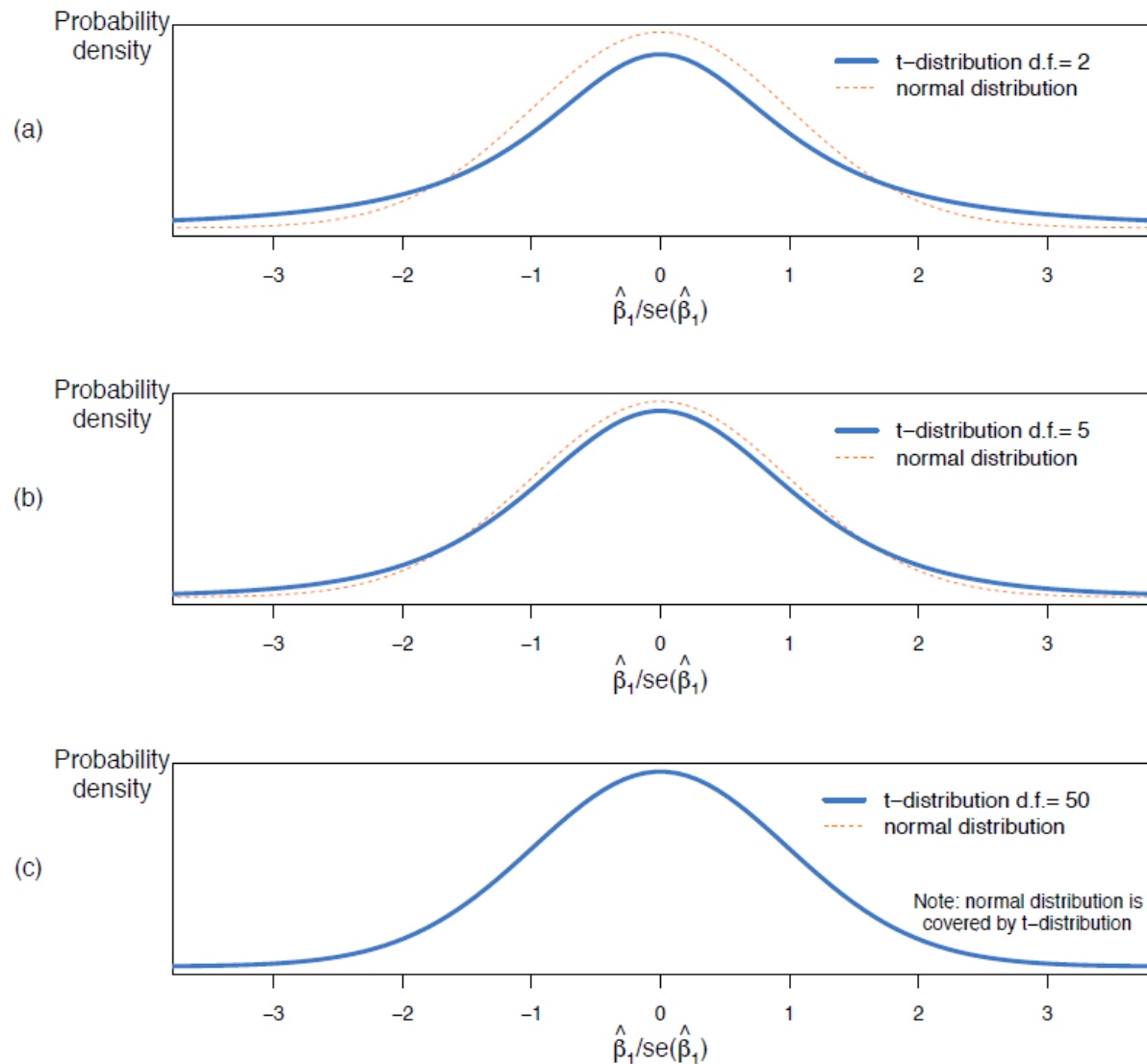
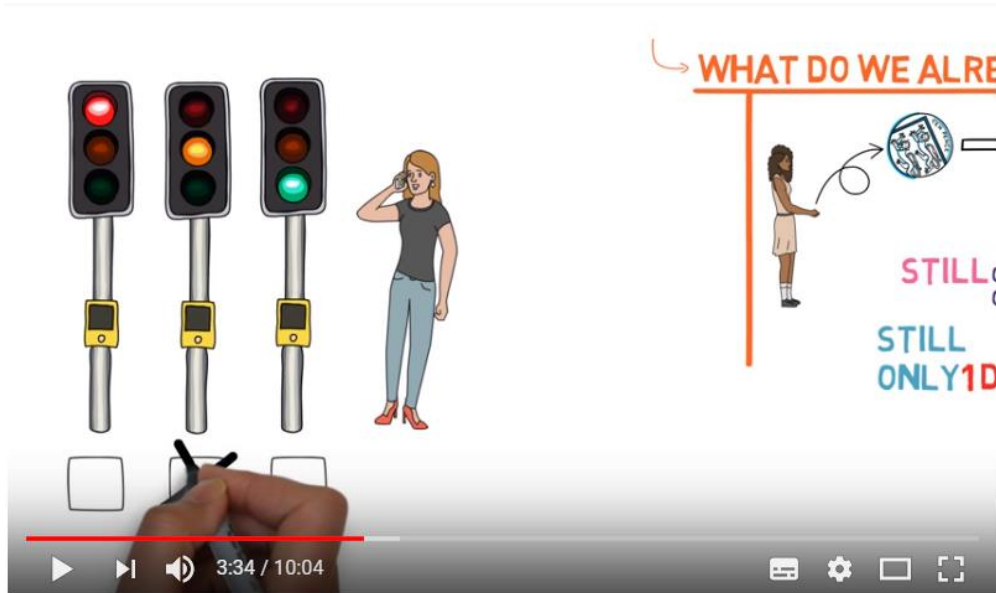


FIGURE 4.3: Three t distributions

Fonte: Bailey (2016: 152).

Como ilustrado na página anterior, a forma específica da distribuição t depende dos graus de liberdade ( $n-k-1$ )



what are degrees of freedom?

<https://youtu.be/rATNoxKg1yA>

**Graus de liberdade:**  
número de observações  
independentes utilizadas  
para calcular uma  
estatística.

# Intuição: Graus de liberdade

Dados o conjunto  $D$  e sua média ( $\bar{D}$ ):

$$D = \{2, 4, 6, 8, 10\}$$

$$\bar{D} = 6$$

Se conheço  $\bar{D}$  e  $N-1$  elementos de  $D$ , então posso determinar o  $n$ -ésimo elemento de  $D$ , sem qualquer dúvida sobre seu valor.

$$\bar{D} = \frac{2+4+6+8+d}{(N-1)+1} = 6$$

$$30 = 20 + d$$

$$d = 30 - 20 = 10$$

Diferente dos elementos até  $N-1$ , o  $n$ -ésimo elemento não é “livre”, pois só poderia admitir um certo valor, dado o que já sei sobre  $D$  (ou seja, dado que conheço  $\bar{D}$ ).



Note que no cálculo de  $\hat{\beta}_1$  precisamos de  $\bar{X}$  e  $\bar{Y}$  para obter os desvios de cada observação em relação à média dessas variáveis. Assim, a média dessas variáveis é conhecida antes de calcularmos o  $\hat{\beta}_1$ . Portanto, a observação  $(X_N, Y_N)$  não é “livre”.

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

O valor crítico é o limiar a partir do qual a estatística  $t$  [ou seja,  $\hat{\beta}_j / \text{se}(\hat{\beta}_j)$ ] é tão improvável que rejeitamos  $H_0$

▪ O valor crítico depende de:

- Graus de liberdade:  $n - k - 1$ ;  
 $k$  = número de variáveis explicativas
- Formato da hipótese alternativa:  $H_1$  unilateral ou bilateral
- Nível de significância estatística:  $\alpha$

Qual é a  $H_0$  em cada um desses painéis?

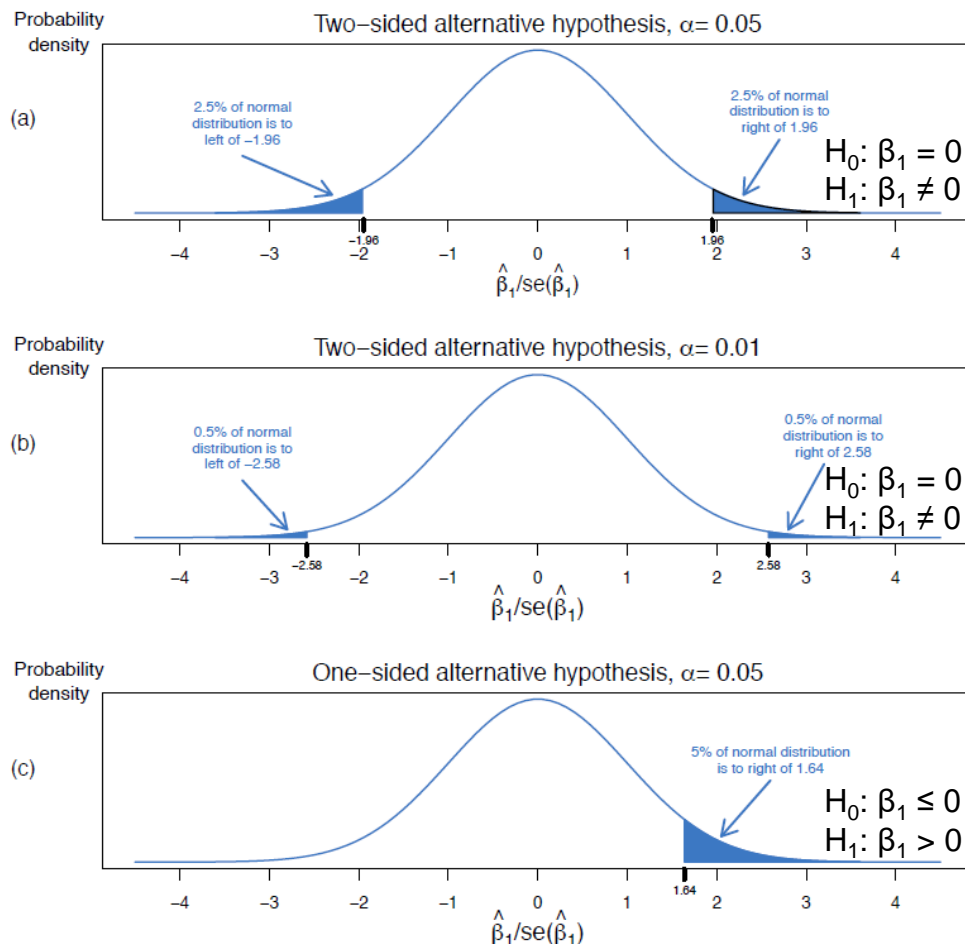
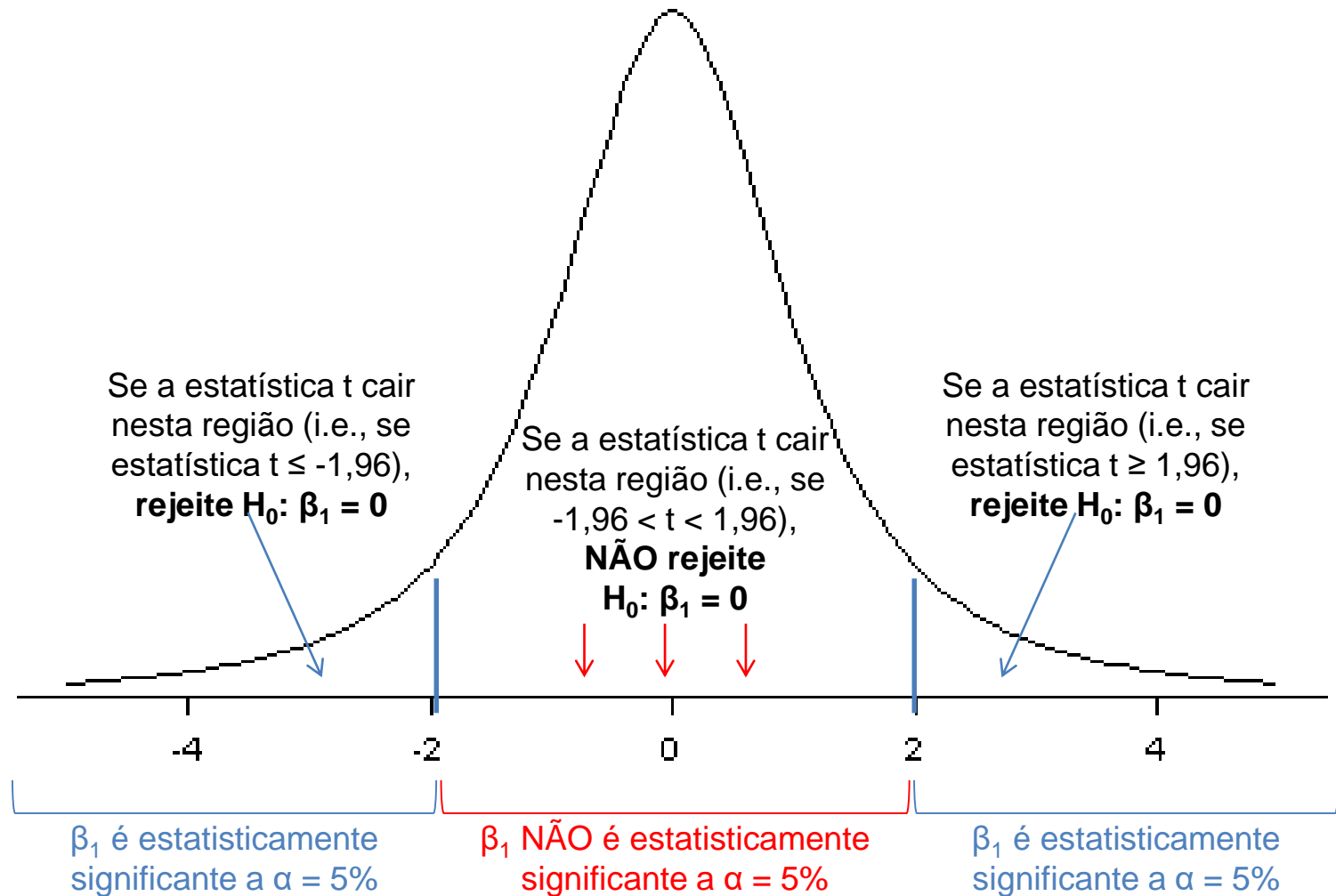


FIGURE 4.4: Critical Values for Large Sample  $t$  tests

Fonte: Bailey (2016: 155).



# Zooming: áreas de rejeição para $H_0: \beta_1 = 0$ (teste bilateral com amostra grande e $\alpha = 0,05$ )



# One- or two-sided?

Better go for two!

Observe que o valor crítico unilateral para  $\alpha = 0,05$  é menor que o valor crítico bilateral. **Os valores críticos unilaterais serão sempre menores para qualquer valor de  $\alpha$ , o que significa que é mais fácil rejeitar a hipótese nula para uma hipótese alternativa unilateral do que para uma hipótese alternativa bilateral.** Portanto, usar valores críticos com base em uma **alternativa bilateral** é **estatisticamente cauteloso** no sentido de que é menos provável que pareçamos ansiosos demais para rejeitar a nula se usarmos uma alternativa bilateral.

Bailey (2016: 156)

Exceto se indicado em contrário, nesta disciplina trataremos de testes t bicaudais.

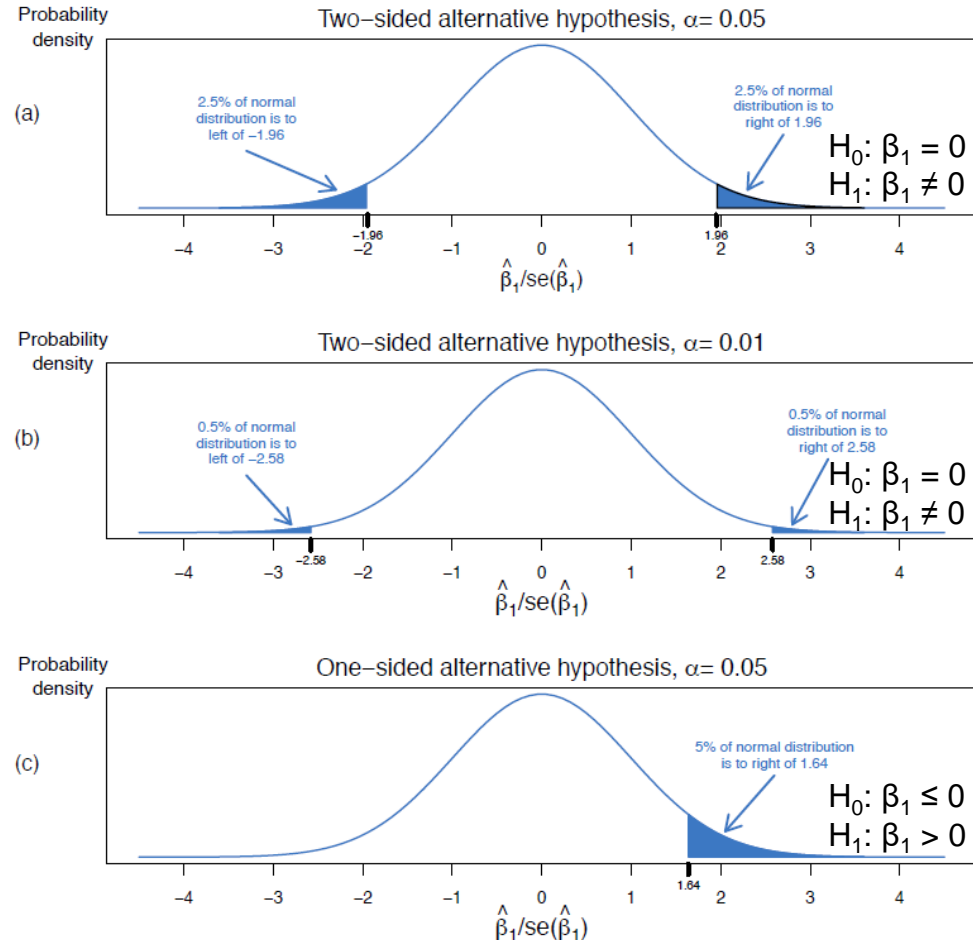


FIGURE 4.4: Critical Values for Large Sample t tests

Fonte: Bailey (2016: 155).

# Valores críticos diminuem à medida que graus de liberdade aumentam

- Quanto **menor** o tamanho da **amostra**, **mais incerteza** temos sobre o  $\beta_j\text{hat}$  e portanto **mais elevado** será o **valor crítico**

[dito de outra forma...]

- Quanto **maior** o tamanho da **amostra**, **menos incerteza** temos sobre o  $\beta_j\text{hat}$  e portanto **menos elevado** será o **valor crítico**

À medida que os graus de liberdade aumentam, a distribuição  $t$  se parece **cada vez mais com uma distribuição normal [padrão]** e, para infinitos graus de liberdade, é exatamente como uma distribuição normal [padrão], produzindo valores críticos idênticos. Para **graus de liberdade acima de 100**, é razoável usar valores críticos da **distribuição normal [padrão]** como uma boa aproximação.

Bailey (2016: 156-157)

Table 4.4: Critical Values for  $t$  distribution

$\alpha$ (1-sided) $\Rightarrow$		0.05	0.025	0.01	0.005
$\alpha$ (2-sided) $\Rightarrow$		0.10	0.050	0.02	0.01
Degrees of freedom	2	2.92	4.30	6.97	9.92
	5	2.01	2.57	3.37	4.03
	10	1.81	2.23	2.76	3.17
	15	1.75	2.13	2.60	2.95
	20	1.73	2.09	2.53	2.85
	50	1.68	2.01	2.40	2.68
	100	1.66	1.98	2.37	2.63
	$\infty$	1.64	1.96	2.32	2.58

A  $t$  distribution with  $\infty$  degrees of freedom is the same as a normal distribution.

Fonte: Bailey (2016: 157).



# DCP098

## Fundamentos para Avaliação Quantitativa de Políticas Públicas

### Teste de hipótese

**Aula 14**  
18 de maio de 2022

Ana Paula Karruz