



DCP098

Fundamentos para Avaliação Quantitativa de Políticas Públicas

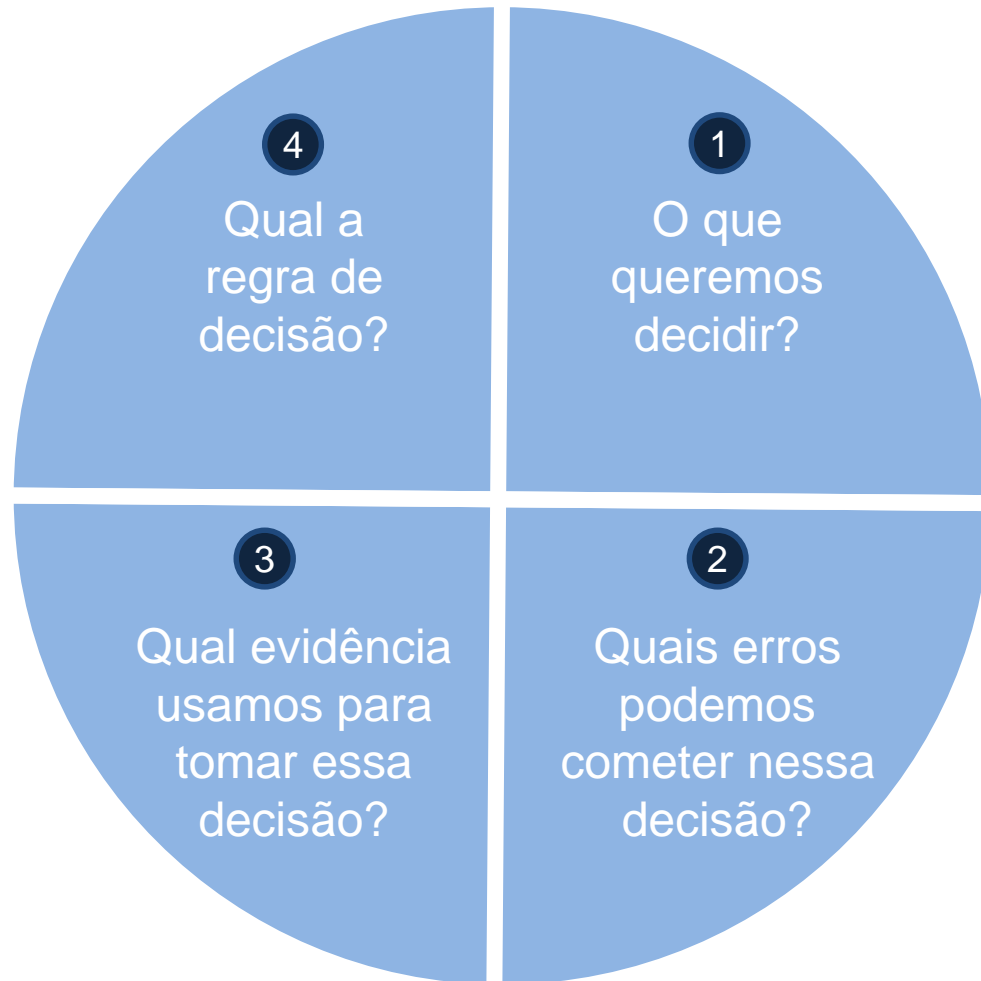
Teste de hipótese

Aulas 15-16
23 e 25 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



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

α (1-sided) \Rightarrow		0.05	0.025	0.01	0.005
α (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
	∞	1.64	1.96	2.32	2.58

A t distribution with ∞ degrees of freedom is the same as a normal distribution.

Fonte: Bailey (2016: 157).

Resumão (até agora)

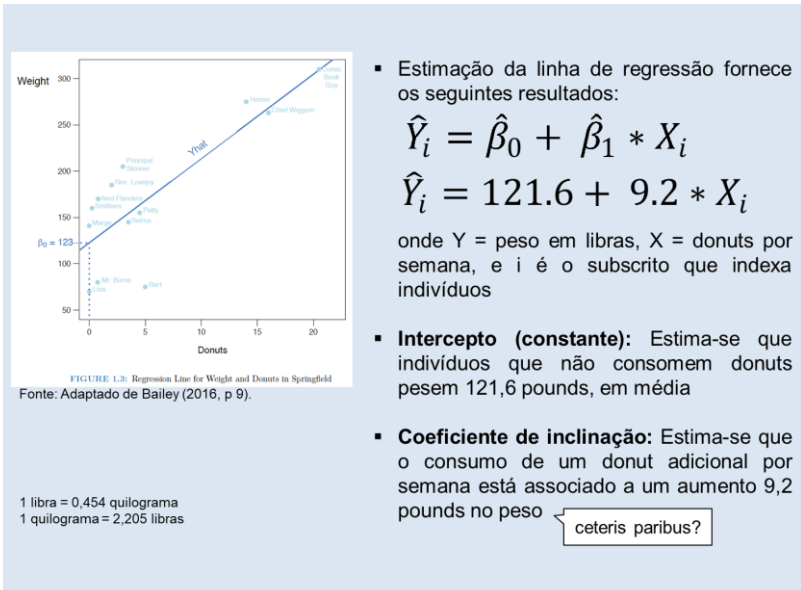
We compare $\frac{\hat{\beta}_1}{se(\hat{\beta}_1)}$ to our critical value and reject H_0 if the magnitude is larger than the critical value. We refer to the ratio of $\frac{\hat{\beta}_1}{se(\hat{\beta}_1)}$ as the **t statistic** (or, “t stat” as the kids say). The t statistic is so named because that ratio will be compared to a critical value that depends on the t distribution in the manner we have just outlined. Tests based on two-sided alternative tests with $\alpha = 0.05$ are very common. When the sample size is large, the critical value for such a test is 1.96. Hence the rule of thumb is that a t statistic bigger than 2 indicates that β_1 is statistically significant at conventional levels.

Fonte: Bailey (2016: 157).

[illegible]

OK, rejeitamos H_0 a 5% de significância.

Qual o menor nível de significância ao qual rejeitaríamos H_0 ?



- Estimação da linha de regressão fornece os seguintes resultados:

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 * X_i$$

$$\hat{Y}_i = 121.6 + 9.2 * X_i$$

onde Y = peso em libras, X = donuts por semana, e i é o subscrito que indexa indivíduos

- Intercepto (constante):** Estima-se que indivíduos que não consomem donuts pesem 121,6 pounds, em média
- Coefficiente de inclinação:** Estima-se que o consumo de um donut adicional por semana está associado a um aumento 9,2 pounds no peso

ceteris paribus?

Num teste bicaudal, p-valor é a probabilidade de se obter uma estatística t de magnitude \geq à estatística t observada se H_0 for verdadeira

- Para β_1 , p-valor = $\Pr(>|t|) = 2 * \Pr(t > |4,707|) = 0,000643$
- Esse é um p-valor bastante baixo, implicando que $\hat{\beta}_1$ (com o grau de imprecisão a ele associado) é realmente muito improvável se H_0 for verdadeira
 - p-valor pequeno fornece evidência contra H_0
 - p-valor alto fornece pouca evidência contra H_0
- Podemos conduzir um teste de hipótese com base no p-valor, aplicando a seguinte regra de decisão:
Se p-valor $\leq \alpha$, então rejeite H_0
Se p-valor $> \alpha$, então não rejeite H_0

$0,000643 < 0,05 \rightarrow$ Rejeite H_0

```
> summary(reg.pounds)
```

```
Call:
```

```
lm(formula = dados$Weight..pounds. ~ dados$Donuts.per.week)
```

```
Residuals:
```

```
      Min       1Q   Median       3Q      Max
-92.731 -13.508   3.916  36.081  55.716
```

```
Coefficients:
```

```
              Estimate Std. Error t value Pr(>|t|)
(Intercept)    121.613    16.593   7.329 1.49e-05 ***
dados$Donuts.per.week  9.224     1.959   4.707 0.000643 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Residual standard error: 45.81 on 11 degrees of freedom
Multiple R-squared:  0.6683, Adjusted R-squared:  0.6381
F-statistic: 22.16 on 1 and 11 DF, p-value: 0.0006426
```

```
> 2*pt(-abs(4.709),df=11)
[1] 0.0006407414
> pt(abs(4.709),df=11)
[1] 0.9996796
```

Além de indicar se rejeitaremos H_0 , o **p-valor informa o nível mínimo de significância ao qual poderíamos rejeitar H_0** ; assim, o p-valor informa o peso da evidência contra H_0 . Neste exemplo, β_1 é estatisticamente significativo (i.e., estatisticamente diferente de zero) a 0,1% (de fato, é significativo a 0,0643%). Portanto, β_1 é também estatisticamente significativo aos níveis convencionais de significância: 1%, 5% e 10% (significância marginal).

Significant other



Amor ou cilada?



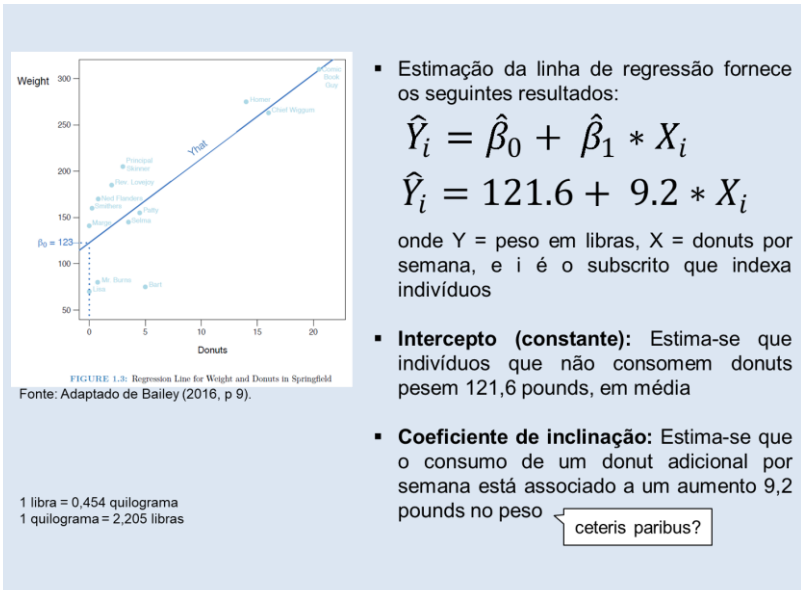
A Statistically Significant Love Song

<https://youtu.be/tVx2V75hWRY>

Uma forma de reconhecer a aleatoriedade de β_j hat é oferecer uma estimativa intervalar de β_j

- **Intuição:** podemos representar nosso grau de confiança sobre o valor do verdadeiro β_j apresentando um intervalo que, acreditamos, contenha esse coeficiente populacional
- **O intervalo de confiança define uma faixa de valores de β_j que seriam os mais consistentes com a evidência (i.e., com β_j hat e seu respectivo grau de imprecisão)**
 - Para os valores de β_j contidos no intervalo de confiança, **nosso β_j hat e respectivo se(β_j hat) não seriam surpreendentes** (i.e., não seriam improváveis)
 - Escolhemos o que entendemos por “improvável” ao definir um nível de significância (α) ou, alternativamente, um nível de confiança ($100\% - \alpha$)
- O intervalo de confiança oferece uma **estimativa intervalar** de β_j , enquanto β_j hat é nossa **estimativa pontual** de β_j

Uma forma de reconhecer a aleatoriedade de β_1 hat é oferecer uma estimativa intervalar de β_1



- Estimação da linha de regressão fornece os seguintes resultados:

$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 * X_i$$

$$\hat{Y}_i = 121.6 + 9.2 * X_i$$

onde Y = peso em libras, X = donuts por semana, e i é o subscrito que indexa indivíduos

- Intercepto (constante):** Estima-se que indivíduos que não consomem donuts pesem 121,6 pounds, em média
- Coefficiente de inclinação:** Estima-se que o consumo de um donut adicional por semana está associado a um aumento 9,2 pounds no peso

ceteris paribus?

```
> summary(reg.pounds)
```

Call:

```
lm(formula = dados$Weight..pounds. ~ dados$Donuts.per.week)
```

Residuals:

Min	1Q	Median	3Q	Max
-92.731	-13.508	3.916	36.081	55.716

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	121.613	16.593	7.329	1.49e-05 ***
dados\$Donuts.per.week	9.224	1.959	4.707	0.000643 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 45.81 on 11 degrees of freedom
Multiple R-squared: 0.6683, Adjusted R-squared: 0.6381
F-statistic: 22.16 on 1 and 11 DF, p-value: 0.0006426

```
> qt(1-0.01/2, 11)
```

```
[1] 3.105807
```

- O intervalo de confiança está **centrado em β_1 hat**
- Sua **extensão** dependente do erro padrão de β_1 hat, dos graus de liberdade e do nível de confiança

$$IC = \beta_1 \text{ hat} \pm [se(\beta_1 \text{ hat}) * t_c]$$

- Para $\alpha = 0,05$ (i.e., nível de confiança = 95%):**
Limite inferior = $9,224 - (1,959 * 2,201) = 4,912$
Limite superior = $9,224 + (1,959 * 2,201) = 13,536$
 $4,912 \leq \beta_1 \leq 13,536$

Interpretação rigorosa: Se repetíssemos o estudo muitas e muitas vezes, com amostras de mesmo tamanho, aproximadamente 95% das estimativas intervalares conteriam o verdadeiro parâmetro populacional

Interpretação coloquial: Estamos 95% confiantes que o verdadeiro β_1 esteja contido no intervalo de confiança

- Para $\alpha = 0,01$ (i.e., nível de confiança = 99%):**
Limite inferior = $9,224 - (1,959 * 3,106) = 3,139$
Limite superior = $9,224 + (1,959 * 3,106) = 15,309$
 $3,139 \leq \beta_1 \leq 15,309$

Quanto maior o nível de confiança, mais largo o intervalo de confiança.

Quais os limites de um IC com 100% de confiança?

Uma forma de reconhecer a aleatoriedade de β_j hat é oferecer uma estimativa intervalar de β_j

Como o **intervalo de confiança** informa a gama de valores de β_j consistentes com a evidência, **também nos indica se rejeitaríamos H_0** .

Se esse **intervalo não incluir zero**, então zero não seria um valor de β_j que provavelmente produziria os dados e as estimativas que observamos e, portanto, **rejeitaríamos $H_0: \beta_j = 0$** , ao nível de significância α .



DEEP DIVE 2:

Qual a intuição por trás do intervalo de confiança?

DEEP DIVE 2:

Qual a intuição por trás do intervalo de confiança? (1/4)

Dada a estimativa $\hat{\beta}$ que obtivemos, quais valores de β seriam consistentes com essa estimativa (i.e., “prováveis”)? Para responder, apoiamo-nos nas afirmações abaixo.

A Na ausência de endogeneidade, $E(\hat{\beta}) = \beta$

Se não houver viés, β é o centro da distribuição de $\hat{\beta}$ e o valor mais provável de $\hat{\beta}$

Implicação: Não sabemos quão próxima de β nossa estimativa se encontra, mas podemos acreditar que provavelmente está próxima, a depender do grau de precisão

B Distribuição teórica das estimativas de $\hat{\beta}$ é simétrica

Se $\hat{\beta}$ for diferente de β , há igual probabilidade de que $\hat{\beta}$ esteja à direita ou à esquerda de β

Implicação: Podemos estimar uma banda de variação para β centrada em $\hat{\beta}$ – até porque não teríamos opção melhor para “ancorar” esse intervalo

C Aleatoriedade na estimação de $\hat{\beta}$ advém tanto de sampling randomness quanto de modeled randomness

Na ausência de endogeneidade, as fontes de incerteza sobre $\hat{\beta}$ são a amostragem (sampling randomness) e a escolha de fatores considerados no modelo explicativo (modeled randomness)

Implicação: Intervalo de confiança deve considerar a aleatoriedade advinda tanto da sampling randomness quanto da modeled randomness

DEEP DIVE 2:

Qual a intuição por trás do intervalo de confiança? (2/4)

Dada a estimativa $\hat{\beta}$ que obtivemos, quais valores de β seriam consistentes com essa estimativa (i.e., “prováveis”)? Para responder, apoiamo-nos nas afirmações abaixo.

D t crítico é uma medida de sampling randomness

Dados os graus de liberdade, t crítico informa uma distância em relação a β^{Nula} ; a probabilidade de valores de $t \geq t$ crítico (ou $t \leq -t$ crítico) é de $(1 - \text{nível de confiança})/2$

Implicações:

- t crítico não é influenciado pela escolha de variáveis explicativas, apenas:
 - Pelo nível de confiança considerado; e
 - Pelos graus de liberdade: $n - k - 1$ (i.e., número de observações livres)
- Portanto, o t crítico reflete apenas um atributo preestabelecido da estimativa intervalar (nível de confiança) e apenas um atributo do modelo (número de graus de liberdade, diretamente afetado pelo tamanho da amostra)

Nota: o t crítico é o mesmo para qualquer $\hat{\beta}$ de um modelo

- t crítico pode ser usado para representar a sampling randomness (quanto maior a amostra, menor a sampling randomness e maiores os graus de liberdade)

DEEP DIVE 2:

Qual a intuição por trás do intervalo de confiança? (3/4)

Dada a estimativa $\hat{\beta}$ que obtivemos, quais valores de β seriam consistentes com essa estimativa (i.e., “prováveis”)? Para responder, apoiamo-nos nas afirmações abaixo.

- E** Erro padrão é uma medida de sampling e modeled randomness combinadas
- O erro padrão informa o grau de imprecisão da distribuição teórica de $\hat{\beta}$; essa imprecisão é causada por sampling e modeled randomness

Implicações:

- O erro padrão é influenciado:
 - Pelo tamanho da amostra: no denominador (no N e variância de X_j)
 - Pelos graus de liberdade: no numerador (na fórmula da variância da regressão);
 - Pela escolha de variáveis explicativas: no numerador (na variância da regressão, que é uma medida de ajuste – quanto maior a variância da regressão, menor o ajuste);
 - Pela variação de X_j : no denominador (na variância de X_j); e
 - [Na regressão múltipla] Pela independência entre variáveis explicativas: no denominador (em $1 - R_j^2$)
- Portanto, o erro padrão reflete os dois tipos de aleatoriedade (tanto sampling quanto modeled randomness)
Nota: O erro padrão é específico ao $\hat{\beta}$ para o qual é estimado
- Erro padrão pode ser usado para representar sampling e modeled randomness

DEEP DIVE 2:

Qual a intuição por trás do intervalo de confiança? (4/4)

Dada a estimativa $\hat{\beta}$ que obtivemos, quais valores de β seriam consistentes com essa estimativa (i.e., “prováveis”)? Para responder, apoiamo-nos nas afirmações abaixo.

F Para corresponder a um certo nível de confiança, intervalo de confiança deve combinar t crítico e erro padrão

Considerando o nível de confiança e os graus de liberdade, t crítico informa sobre a região onde β é provável de estar (dado o $\hat{\beta}$ estimado), mas é “genérico”: para um certo nível de confiança, será o mesmo para qualquer modelo com os mesmos graus de liberdade

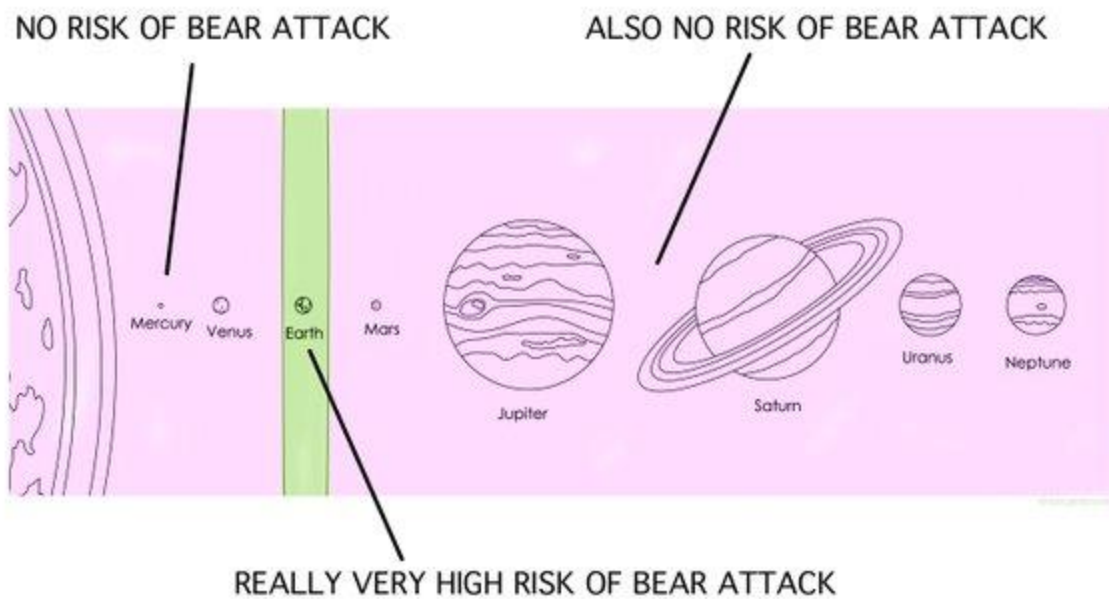
A seu turno, erro padrão reflete sampling e modeled randomness, mas nada diz sobre a região onde β é mais provável de estar (dado o $\hat{\beta}$ estimado)

Implicações:

- Intervalo de confiança não pode se apoiar unicamente no t crítico ou no erro padrão, devendo combinar essas medidas de aleatoriedade
- A um certo nível de confiança, ao se multiplicar t crítico pelo erro padrão:
 - Se erro padrão for igual à unidade, o intervalo de confiança é determinado pelo t crítico
 - Se erro padrão for inferior à unidade, estreita-se o intervalo de confiança
 - Se erro padrão for superior à unidade, dilata-se o intervalo de confiança



CHART TO HELP DETERMINE RISK OF BEAR ATTACK:



(Very accurate, but not precise.)

Testes de hipótese apoiam a inferência estatística (extração de conclusões a partir da amostra), mas têm limitações

Testes de significância estatística...

- Nada têm a ver com a **solidez teórica** do modelo
- São baseados na premissa de que **não há endogeneidade**
- Podem conduzir a **diferentes conclusões para cenários virtualmente iguais**: $|t|$ apenas ligeiramente superior (inferior) ao t crítico
- Podem sugerir **conclusões idênticas para cenários dramaticamente diferentes**: $|t|$ muito (apenas ligeiramente) superior ao t crítico
- São afetados pelo tamanho da amostra: **quanto maior a amostra, maior a chance de encontrar significância estatística**. Isto ocorre porque o t crítico é menor para amostras maiores; também, o erro padrão tende a ser menor quanto maior for o tamanho da amostra
 - Analogamente, em **amostras pequenas**, o erro padrão tende a ser alto, de forma que o teste pode **não rejeitar H_0** , ainda que β_j tenha magnitude expressiva
- Não indicam **significância material**: β_j pode ser estatisticamente significativa ainda que sua magnitude seja desprezível em termos práticos

p-valor
ajuda
aqui

Para saber mais sobre p-valor e limites dos testes de hipótese



Figueiredo Filho et al. (2013)



Figueiredo Filho et al. (2014)



Silva e Guarnieri (2014)

Downloaded by [University of Connecticut] at 22:40 29 August 2017

THE AMERICAN STATISTICIAN
2016, VOL. 70, NO. 2, 129–133
<http://dx.doi.org/10.1080/00031305.2016.1154108>



EDITORIAL

The ASA's Statement on p-Values: Context, Process, and Purpose

In February 2014, George Cobb, Professor Emeritus of Mathematics and Statistics at Mount Holyoke College, posed these questions to an ASA discussion forum:

- Q: Why do so many colleges and grad schools teach $p = 0.05$?
A: Because that's still what the scientific community and journal editors use.
Q: Why do so many people still use $p = 0.05$?
A: Because that's what they were taught in college or grad school.

Cobb's concern was a long-worrisome circularity in the sociology of science based on the use of bright lines such as $p < 0.05$: "We teach it because it's what we do; we do it because it's what we teach." This concern was brought to the attention of the ASA Board.

The ASA Board was also stimulated by highly visible discussions over the last few years. For example, ScienceNews (Siegfried 2010) wrote: "It's science's dirtiest secret: The 'scientific method' of testing hypotheses by statistical analysis stands on a flimsy foundation." A November 2013, article in Phys.org Science News Wire (2013) cited "numerous deep flaws" in null hypothesis significance testing. A ScienceNews article (Siegfried 2014) on February 7, 2014, said "statistical techniques for testing hypotheses ... have more flaws than Facebook's privacy policies." A week later, statistician and "Simply Statistics" blogger Jeff Leek responded, "The problem is not that people use P-values poorly," Leek wrote, "it is that the vast majority of data analysis is not performed by people properly trained to perform data analysis" (Leek 2014). That same week, statistician and science writer Regina Nuzzo published an article in *Nature* entitled "Scientific Method: Statistical Errors" (Nuzzo 2014). That article is now one of the most highly viewed *Nature* articles, as reported by altmetric.com (<http://www.altmetric.com/details/2115792#score>).

Of course, it was not simply a matter of responding to some articles in print. The statistical community has been deeply concerned about issues of *reproducibility* and *replicability* of scientific conclusions. Without getting into definitions and distinctions of these terms, we observe that much confusion and even doubt about the validity of science is arising. Such doubt can lead to radical choices, such as the one taken by the editors of *Basic and Applied Social Psychology*, who decided to ban p-values (null hypothesis significance testing) (Trafimow and Marks 2015). Misunderstanding or misuse of statistical inference is only one cause of the "reproducibility crisis" (Peng 2015), but to our community, it is an important one.

When the ASA Board decided to take up the challenge of developing a policy statement on p-values and statistical significance, it did so recognizing this was not a lightly taken step. The ASA has not previously taken positions on specific matters of statistical practice. The closest the association has come to this is a statement on the use of value-added models (VAM) for educational assessment (Morganstein and Wasserstein

2014) and a statement on risk-limiting post-election audits (American Statistical Association 2010). However, these were truly policy-related statements. The VAM statement addressed a key educational policy issue, acknowledging the complexity of the issues involved, citing limitations of VAMs as effective performance models, and urging that they be developed and interpreted with the involvement of statisticians. The statement on election auditing was also in response to a major but specific policy issue (close elections in 2008), and said that statistically based election audits should become a routine part of election processes.

By contrast, the Board envisioned that the ASA statement on p-values and statistical significance would shed light on an aspect of our field that is too often misunderstood and misused in the broader research community, and, in the process, provides the community a service. The intended audience would be researchers, practitioners, and science writers who are not primarily statisticians. Thus, this statement would be quite different from anything previously attempted.

The Board tasked Wasserstein with assembling a group of experts representing a wide variety of points of view. On behalf of the Board, he reached out to more than two dozen such people, all of whom said they would be happy to be involved. Several expressed doubt about whether agreement could be reached, but those who did said, in effect, that if there was going to be a discussion, they wanted to be involved.

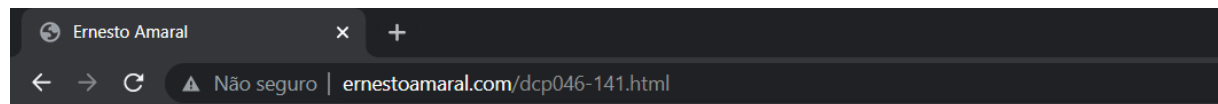
Over the course of many months, group members discussed what format the statement should take, tried to more concretely visualize the audience for the statement, and began to find points of agreement. That turned out to be relatively easy to do, but it was just as easy to find points of intense disagreement.

The time came for the group to sit down together to hash out these points, and so in October 2015, 20 members of the group met at the ASA Office in Alexandria, Virginia. The 2-day meeting was facilitated by Regina Nuzzo, and by the end of the meeting, a good set of points around which the statement could be built was developed.

The next 3 months saw multiple drafts of the statement, reviewed by group members, by Board members (in a lengthy discussion at the November 2015 ASA Board meeting), and by members of the target audience. Finally, on January 29, 2016, the Executive Committee of the ASA approved the statement.

The statement development process was lengthier and more controversial than anticipated. For example, there was considerable discussion about how best to address the issue of multiple potential comparisons (Gelman and Loken 2014). We debated at some length the issues behind the words "a p-value near 0.05 taken by itself offers only weak evidence against the null

ANEXO



[Home](#)

ERNESTO F. L. AMARAL

**Departamento de Ciência Política (DCP)
Faculdade de Filosofia e Ciências Humanas (FAFICH)
Universidade Federal de Minas Gerais (UFMG)**

(2014-1b) DCP 046 – Avaliação de Políticas Públicas

Período: 1º semestre de 2014

Horário: terças-feiras (20:50 às 22:30) e sextas-feiras (19:00 às 20:40)

Aulas teóricas: FAFICH 3011

Aulas práticas: FAFICH 3062

Atendimento aos alunos: FAFICH 209, terças-feiras (14:00 às 19:00), com
marcação por email

Carga horária: 60 horas/aula (4 créditos)

Professor:

Ernesto Friedrich de Lima Amaral

Email: eflamaral@gmail.com

<http://www.ernestoamaral.com/dcp046-141.html>

Tipos de teste t: Slides do Prof. Ernesto Amaral

(DCP134 – Avaliação de Políticas Públicas, 2014a, Aula 20-21)

8

TESTE: HIPÓTESES ALTERNATIVAS UNILATERAIS

$$H_1: \beta_j > 0 \quad \text{OU} \quad H_1: \beta_j < 0$$

- Devemos decidir sobre um nível de significância (geralmente de 5%).
- Corremos o risco de rejeitar erroneamente H_0 , quando ela é verdadeira, em 5% das vezes (erro tipo I igual ao α).
- Um valor suficientemente grande de t , com um nível de significância de 5%, é o 95º percentil de uma distribuição t com $n-k-1$ graus de liberdade (ponto c).
- **Regra de rejeição** é que H_0 é rejeitada em favor de H_1 , se $t > c$ ($H_1: \beta_j > 0$) ou $t < -c$ ($H_1: \beta_j < 0$), em um nível específico.
- Quando os graus de liberdade da distribuição t ficam maiores, a distribuição t aproxima-se da distribuição normal padronizada.
- Para graus de liberdade maiores que 120, pode-se usar os valores críticos da distribuição normal padronizada.

Tipos de teste t: Slides do Prof. Ernesto Amaral

(DCP134 – Avaliação de Políticas Públicas, 2014a, Aula 20-21)

9

REGRA DE REJEIÇÃO DE H_0 (UNILATERAL)

Figura 4.2

Regra de rejeição a 5% para a hipótese alternativa $H_1: \beta_j > 0$ com 28 gl.

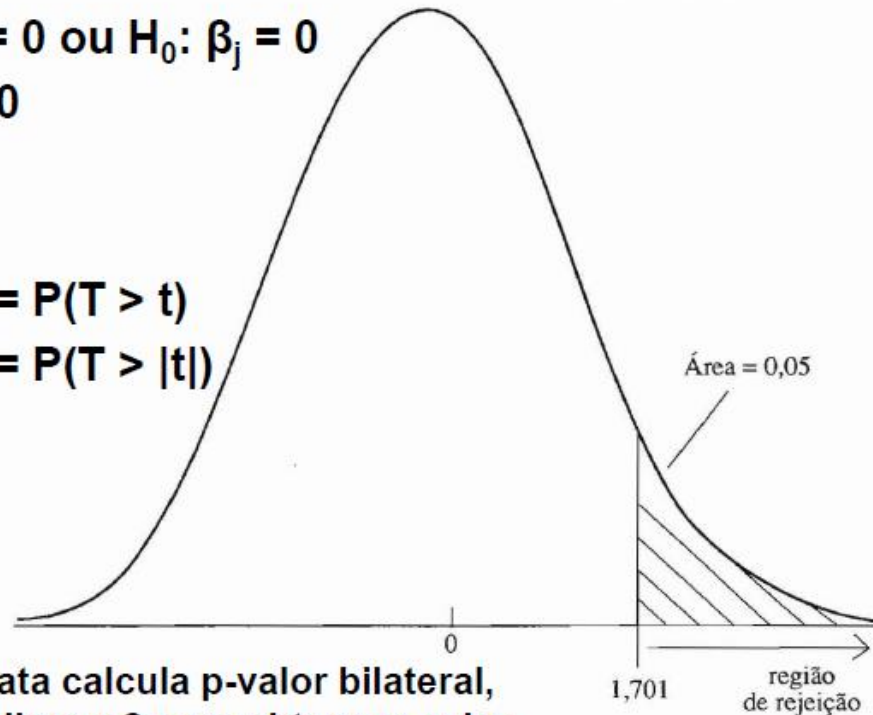
$$H_0: \beta_j \leq 0 \text{ ou } H_0: \beta_j = 0$$

$$H_1: \beta_j > 0$$

$$t_{\beta_j} > c$$

$$\text{p-valor} = P(T > t)$$

$$\text{p-valor} = P(T > |t|)$$



Como Stata calcula p-valor bilateral, é só dividir por 2 para obter o p-valor unilateral.

Fonte: Wooldridge, 2008: 117.

Tipos de teste t: Slides do Prof. Ernesto Amaral

(DCP134 – Avaliação de Políticas Públicas, 2014a, Aula 20-21)

10

REGRA DE REJEIÇÃO DE H_0 (UNILATERAL)

Figura 4.3

Regra de rejeição a 5% para a hipótese alternativa $H_1: \beta_j < 0$, com 18 gl.

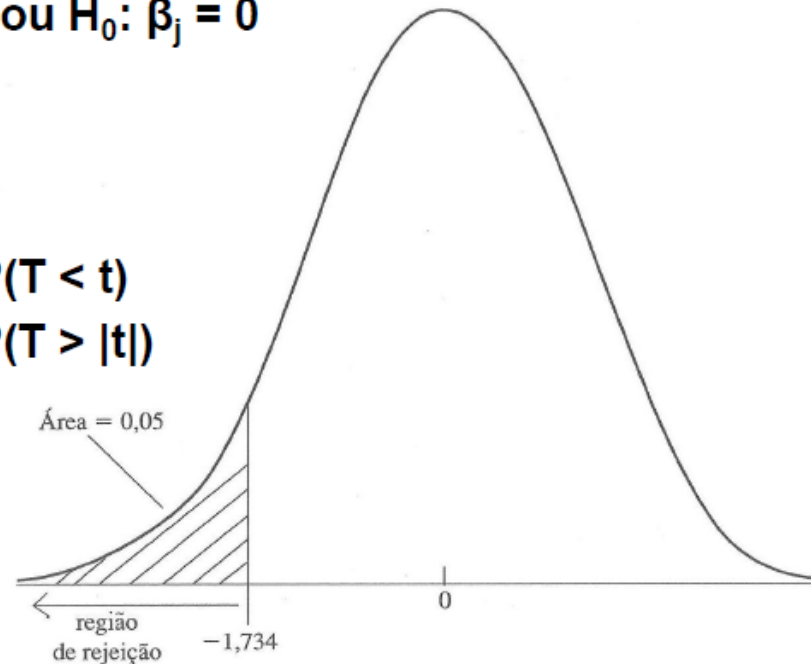
$$H_0: \beta_j \geq 0 \text{ ou } H_0: \beta_j = 0$$

$$H_1: \beta_j < 0$$

$$t_{\beta_j} < -c$$

$$\text{p-valor} = P(T < t)$$

$$\text{p-valor} = P(T > |t|)$$



Como Stata calcula p-valor bilateral, é só dividir por 2 para obter o p-valor unilateral.

Fonte: Wooldridge, 2008: 119.

Tipos de teste t: Slides do Prof. Ernesto Amaral

(DCP134 – Avaliação de Políticas Públicas, 2014a, Aula 20-21)

11

TESTE: HIPÓTESES ALTERNATIVAS BILATERAIS

$$H_1: \beta_j \neq 0$$

- Essa hipótese é relevante quando o sinal de β_j não é bem determinado pela teoria.
- Usar as estimativas da regressão para nos ajudar a formular as hipóteses nula e alternativa não é permitido, porque a inferência estatística clássica pressupõe que formulamos as hipóteses nula e alternativa sobre a população antes de olhar os dados.
- Quando a alternativa é bilateral, estamos interessados no valor absoluto da estatística t : $|t| > c$.
- Para um nível de significância de 5% e em um teste bi-caudal, c é escolhido de forma que a área em cada cauda da distribuição t seja igual a 2,5%.
- Se H_0 é rejeitada, x_j é estatisticamente significativa (ou estatisticamente diferente de zero) ao nível de 5%.

Tipos de teste t: Slides do Prof. Ernesto Amaral

(DCP134 – Avaliação de Políticas Públicas, 2014a, Aula 20-21)

12

REGRA DE REJEIÇÃO DE H_0 (BILATERAL)

Figura 4.4

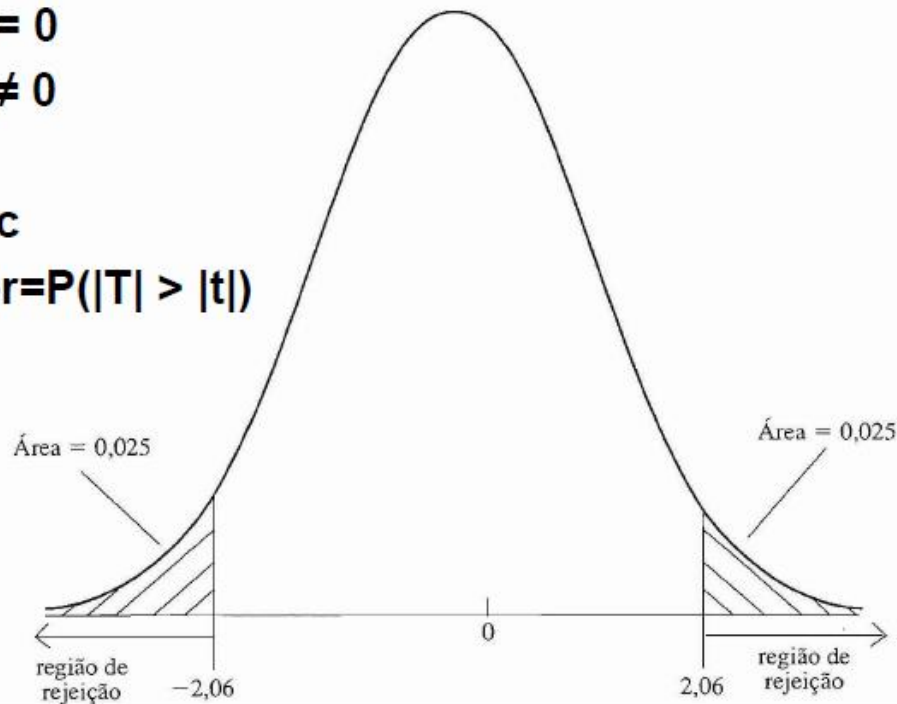
Regra de rejeição a 5% para a hipótese alternativa $H_1: \beta_j \neq 0$ com 25 gl.

$$H_0: \beta_j = 0$$

$$H_1: \beta_j \neq 0$$

$$|t_{\beta_j}| > c$$

$$\text{p-valor} = P(|T| > |t|)$$



Fonte: Wooldridge, 2008: 122.

Tipos de teste t: Slides do Prof. Ernesto Amaral

(DCP134 – Avaliação de Políticas Públicas, 2014a, Aula 20-21)

13

EXEMPLO DE NÃO-REJEIÇÃO DE H_0 (BILATERAL)

Figura 4.6

Obtendo o p -valor contra uma alternativa bilateral, quando $t = 1,85$ e $gl = 40$.

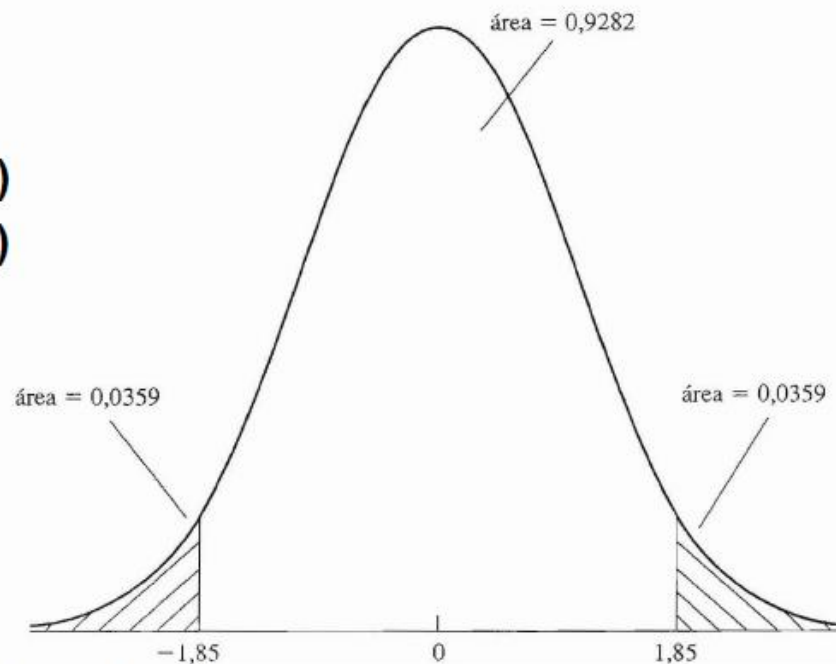
p-valor

$$\begin{aligned} &= P(|T| > |t|) \\ &= P(|T| > 1,85) \\ &= 2P(T > 1,85) \\ &= 2(0,0359) \\ &= 0,0718 \end{aligned}$$

p-valor $> \alpha$

$$0,0718 > 0,05$$

$H_0 : \beta_j = 0$ não é rejeitada



Fonte: Wooldridge, 2008: 127.

Tipos de teste t: Slides do Prof. Ernesto Amaral

(DCP134 – Avaliação de Políticas Públicas, 2014a, Aula 20-21)

14

TESTES DE OUTRAS HIPÓTESES SOBRE β_j

- Poderíamos supor que uma variável dependente (log do número de crimes) necessariamente será relacionada positivamente com uma variável independente (log do número de estudantes matriculados na universidade).
- A hipótese alternativa testará se o aumento de 1% nas matrículas aumentará o crime em mais de 1%:

$$H_0: \beta_j = 1$$

$$H_1: \beta_j > 1$$

- $t = (\text{estimativa} - \text{valor hipotético}) / (\text{erro-padrão})$
- Neste exemplo, $t = (\beta_j - 1) / \text{ep}(\beta_j)$
- Observe que adicionar 1 na hipótese nula, significa subtrair 1 no teste t .
- Rejeitamos H_0 se $t > c$, em que c é o valor crítico unilateral.



DCP098

Fundamentos para Avaliação Quantitativa de Políticas Públicas

Teste de hipótese

Aulas 15-16
23 e 25 de maio de 2022

Ana Paula Karruz