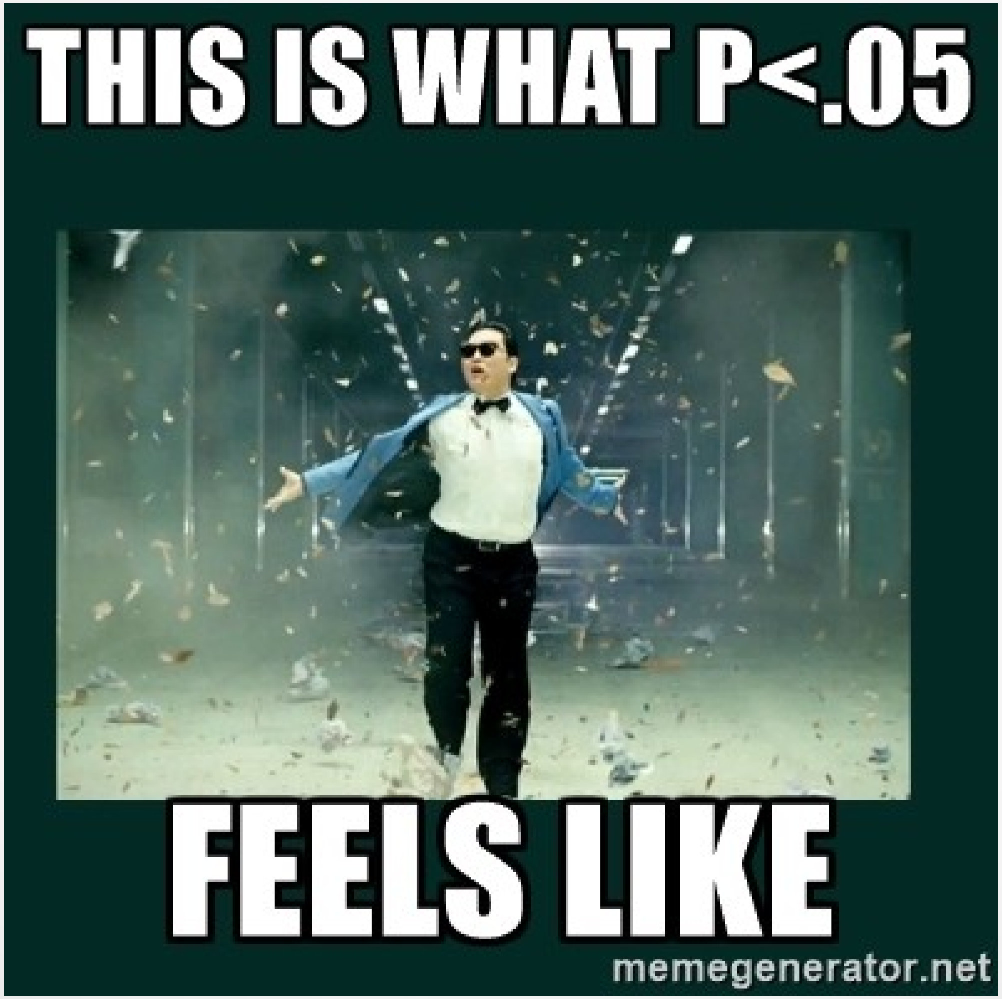
# P<0.05?

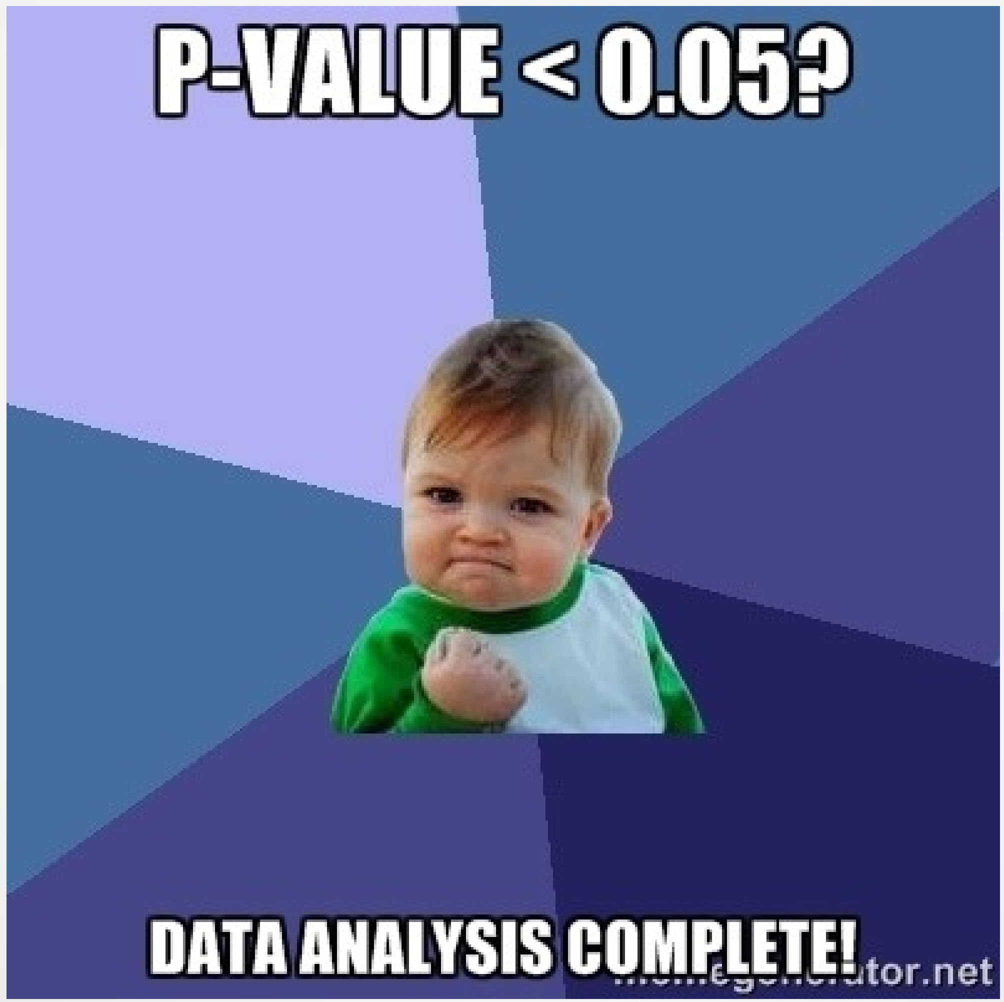
Mitos e verdade sobre o valor-p

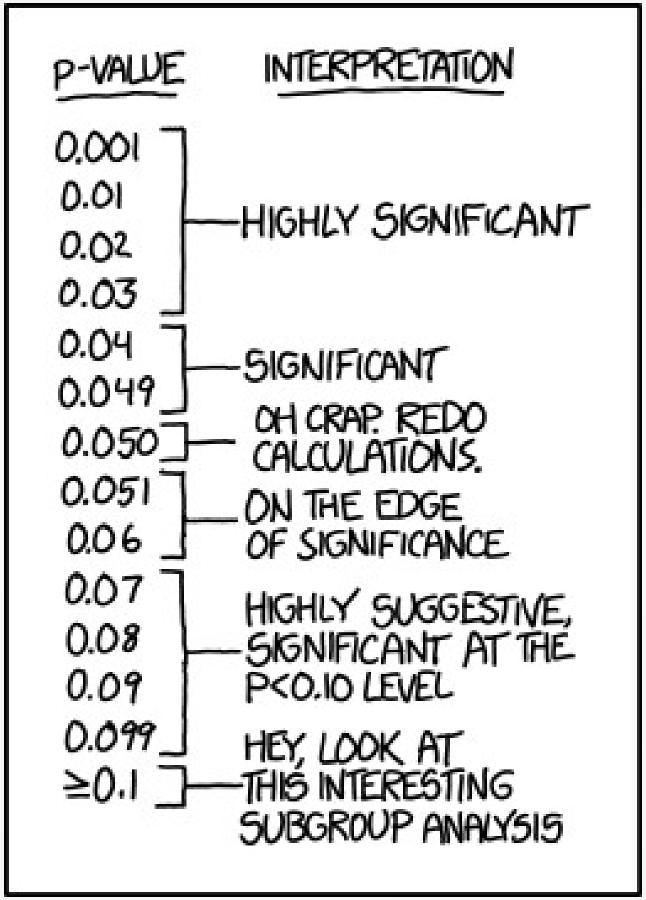
Leo Bastos

PROCC/Fiocruz

Contato: leonardo.bastos@fiocruz.br


THIS IS WHAT P<.05
FEELS LIKE
memegenerator.net

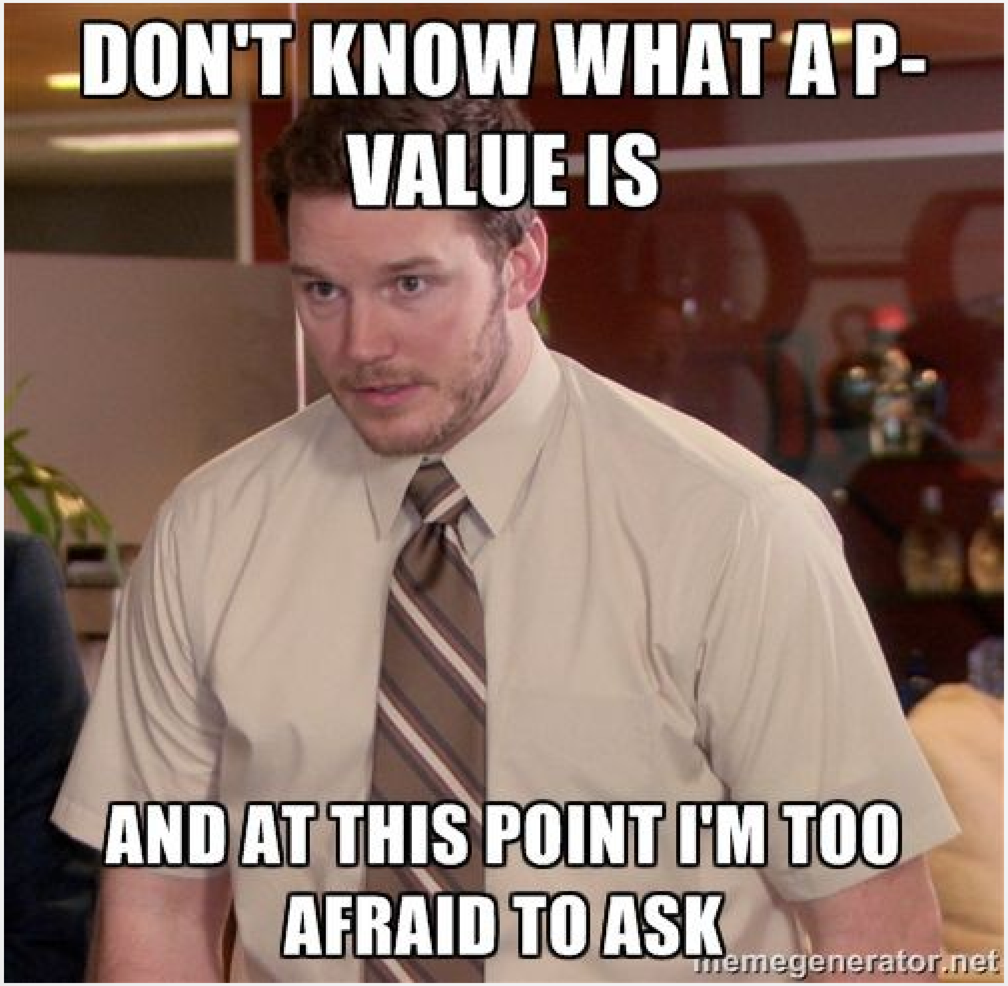

P-VALUE < 0.05?
DATA ANALYSIS COMPLETE!



Source: xkcd.org

Para que serve o valor-p?

- Extremamente usado para mostrar significância estatística nas mais diversas situações:
  - Comparação de dois grupos (controle x tratamento)
    - Teste-Z, teste-t, Wilcoxon,...
  - Testes de associação
    - Teste chi-quadrado, Exato de Fisher
  - Testes de aderência
    - Anderson-Darling, Kolmogorv-Smirnov, chi-quadrado
  - ANOVA
  - Testes baseados no TCL, normalidade assintótica EMV


DON'T KNOW WHAT A P-VALUE IS
AND AT THIS POINT I'M TOO AFRAID TO ASK
memegenerator.net

Um pouquinho de história

- A teoria de testes de hipóteses foi desenvolvida por três* pesquisadores, nas décadas de 1920 e 1930

Karl Pearson (1857 – 1936)

Ronald Fisher (1890 – 1961)

Jerzy Neyman (1894 – 1981)

Egon Pearson (1895-1980)

Um pouquinho de história

- Fisher desenvolveu a modelagem estatística (Fisher, 1922).
  - O valor-p deveria ser visto como uma medida contínua. Valores muito pequenos dessa estatística de teste mostram o quão improváveis seus dados vieram do modelo proposto.

- Neyman e Pearson formalizaram o processo de tomada de decisão via teste de hipóteses. (Neyman & Pearson, 1933)
  - O valor-p então deveria ser comparado com um ponto de corte que **você** deveria julgar como erro aceitável a ser cometido em sua pesquisa (erro tipo I).

Teste de hipóteses e tipos de erro

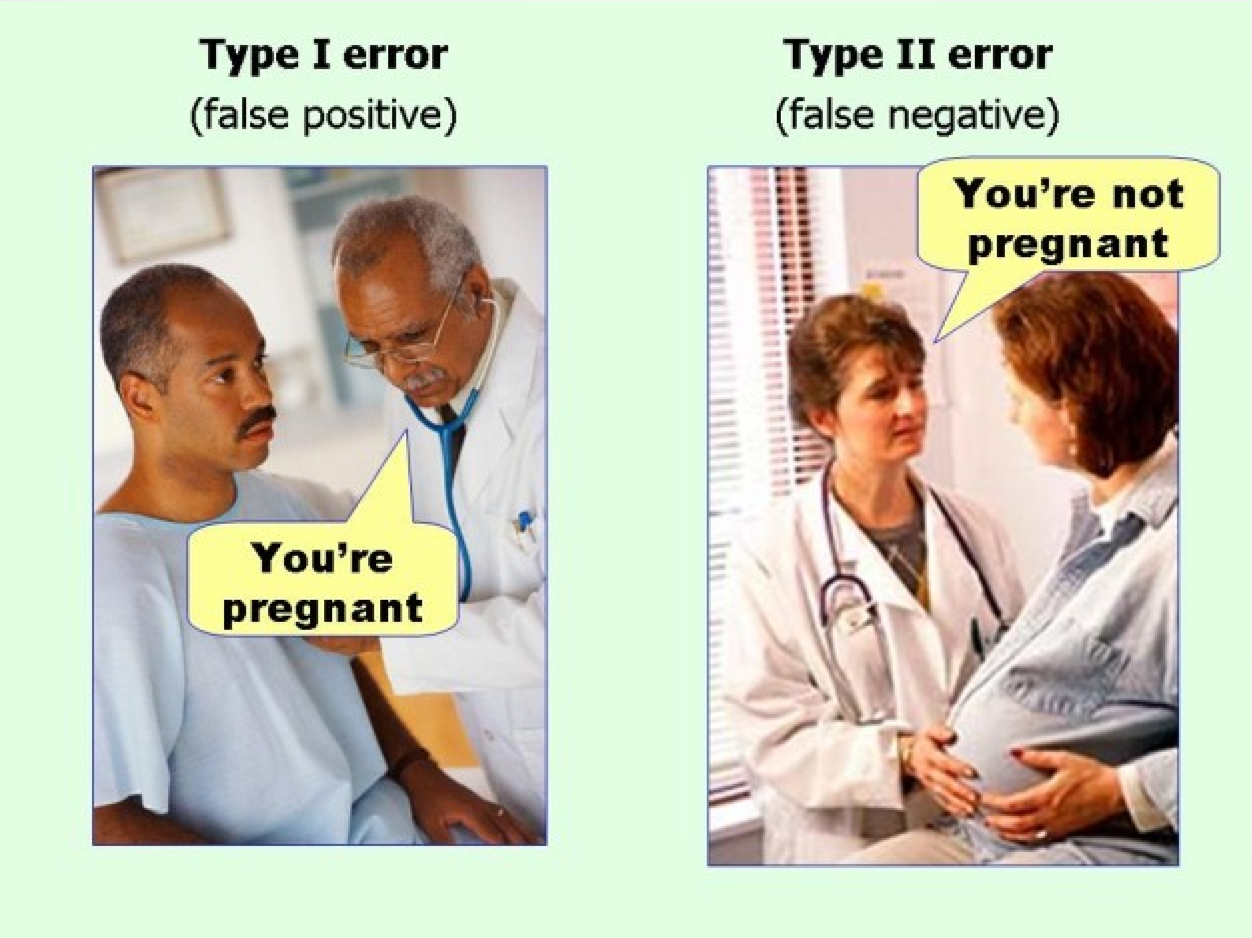
H0: Condição default
H1: usualmente a novidade a ser testada

Decisão	H0 é verdadeira	H1 é verdadeira
Aceitar H0	Decisão correta	Erro tipo II
Rejeitar H0	Erro tipo I	Decisão correta

Os testes de hipóteses estatisticos geralmente são construídos de forma que ao fixarmos um erro (usualmente erro tipo I), o outro erro será o menor possível. (testes UMP).

O uso do valor-p nos ajuda a tomar uma melhor decisão.

Tipos de erro



O que o valor-p NÃO é?

- Antes de dizer o que é, vamos ver o que ele NÃO é:
  - NÃO é a probabilidade da hipótese nula ser verdadeira;
  - NÃO é a probabilidade de que os dados foram gerados totalmente ao acaso;
  - NÃO representa as chances do modelo nulo ser verdadeiro;
  - NÃO é o erro tipo I, ou tamanho do teste;
  - NÃO é o poder do teste;

O que é o valor-p?

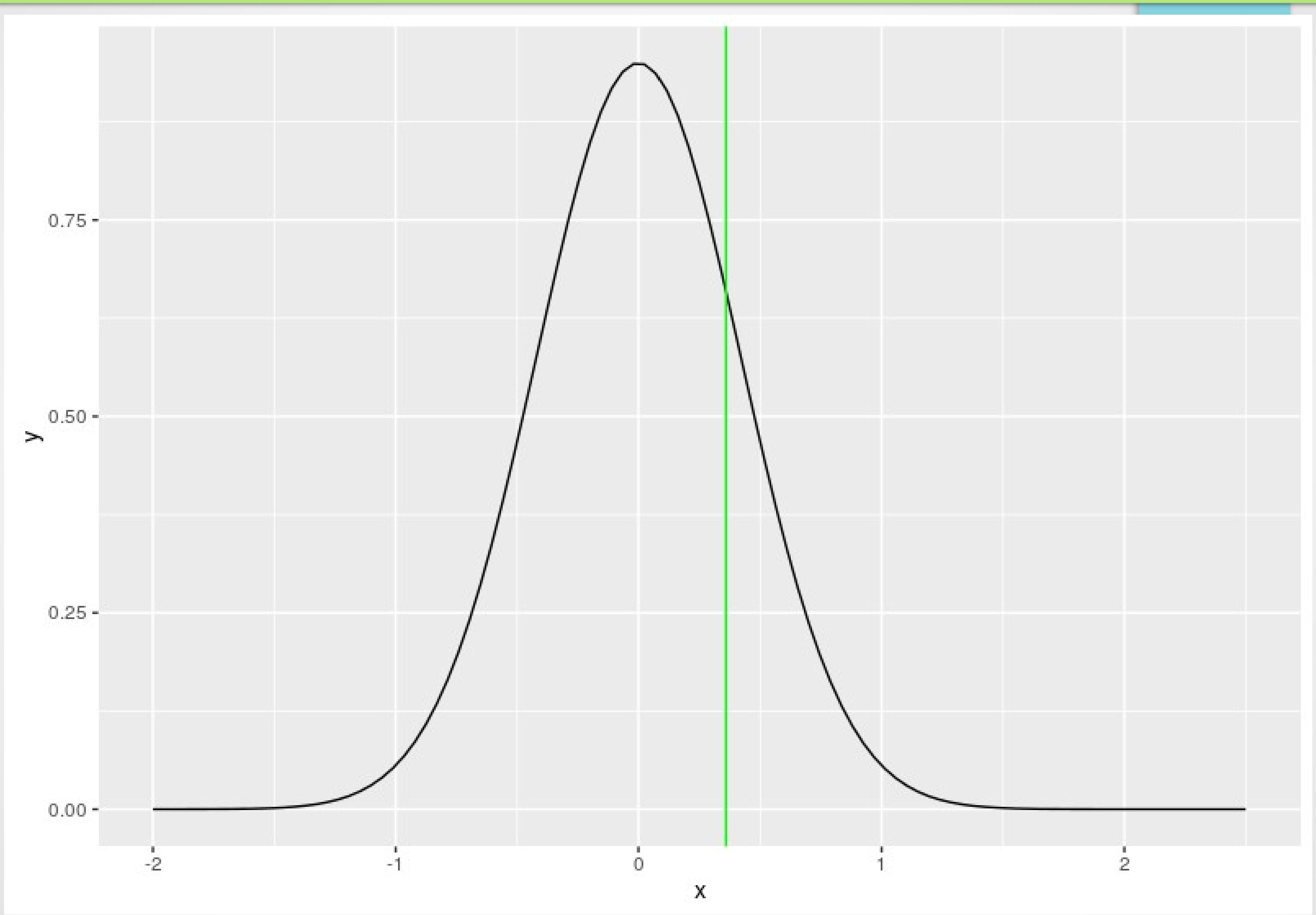
- **O valor-p é a probabilidade, sob o modelo estatístico especificado em H0, de um resumo estatístico dos dados fornecer resultados iguais ou mais extremos que o observado.**
- Outras definições:
  - P-valor é o menor nível de significância para o qual a hipótese nula seria rejeitada.
  - P-valor é uma estatística de teste específica onde pequenos valores fornecem evidências contra a hipótese nula.

Exemplo teste-t

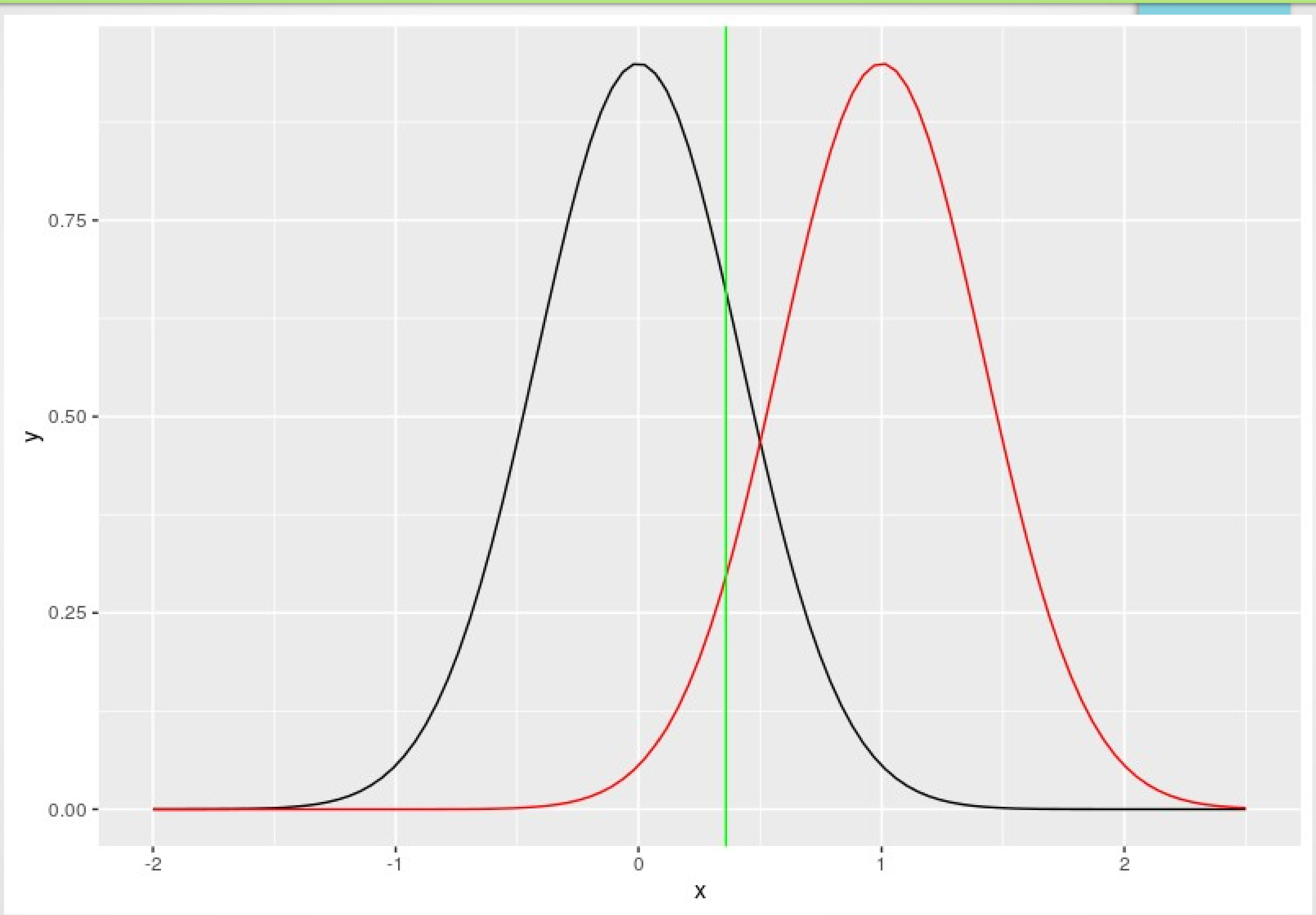
```r
x <- rnorm(50, mean = 1, sd = 3)

t.test(x, mu = 0)

##
##  One Sample t-test
##
## data:  x
## t = 0.9565, df = 49, p-value = 0.3435
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  -0.3954207   1.1137388
## sample estimates:
## mean of x
## 0.359159
```

$$\begin{cases} H_0: & \mu = 0 \\ H_A: & \mu \neq 0 \end{cases}$$

Exemplo teste-t (Sob H0: mu = 0)



Exemplo teste-t (Sob H0: mu = 0)



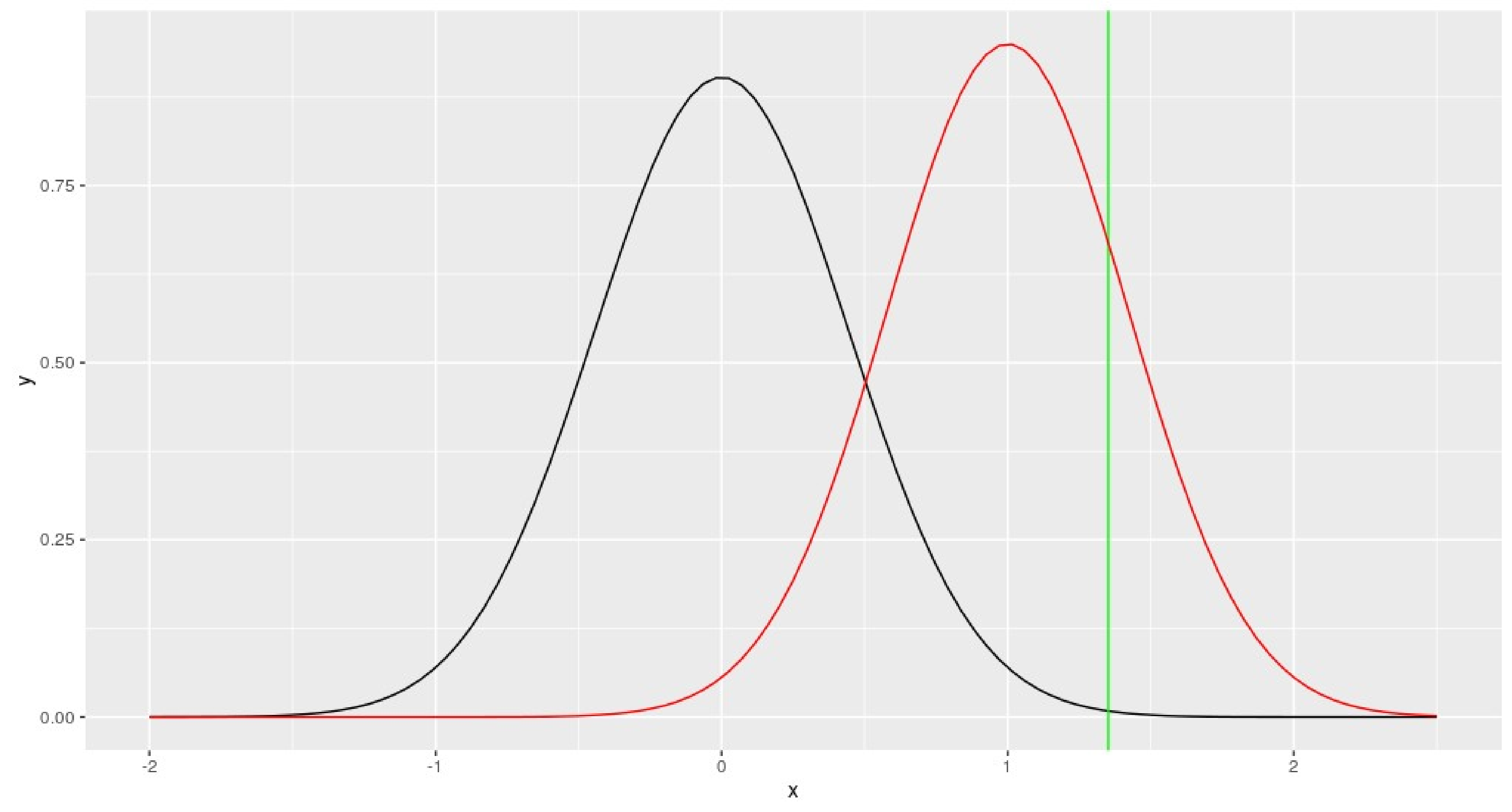
Exemplo teste-t: Outra amostra

```r
x <- rnorm(50, 1, sd = 3)

t.test(x, mu = 0)
```

$$\begin{cases} H_0: & \mu = 0 \\ H_A: & \mu \neq 0 \end{cases}$$

```
##
##  One Sample t-test
##
## data:  x
## t = 3.06, df = 49, p-value = 0.003585
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  0.4641282 2.2400470
## sample estimates:
## mean of x
##  1.352088
```

Exemplo teste-t: Outra amostra



False discovery rate

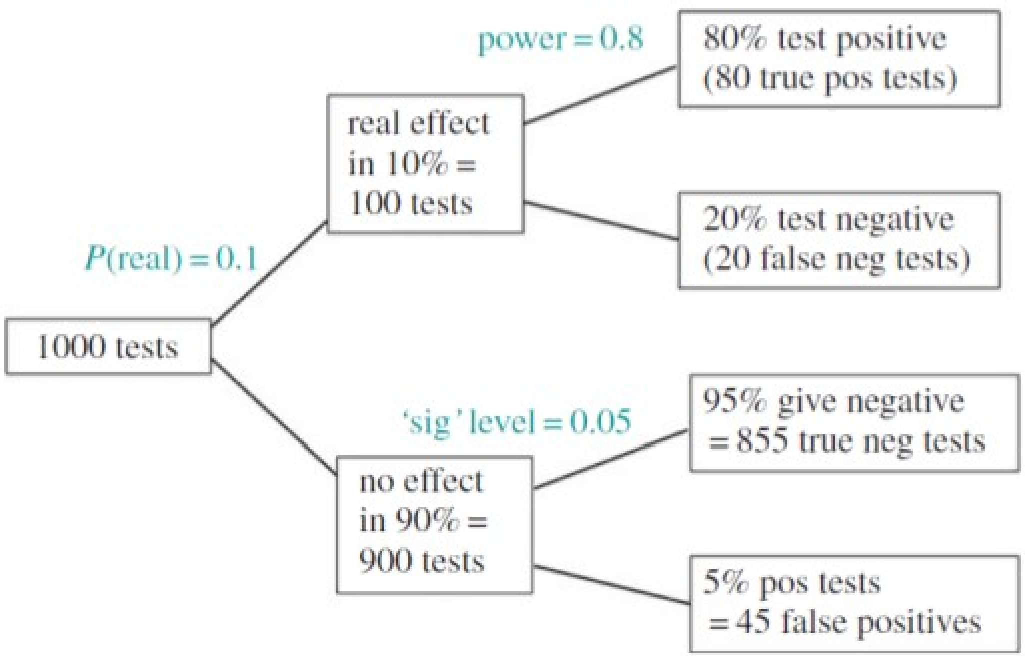


Figure 2. Tree diagram to illustrate the false discovery rate in significance tests. This example considers 1000 tests, in which the prevalence of real effects is 10%. The lower limb shows that with the conventional significance level, $p = 0.05$, there will be 45 false positives. The upper limb shows that there will be 80 true positive tests. The false discovery rate is therefore $45/(45 + 80) = 36\%$, far bigger than 5%.

False discovery rate

P(real)	False discovery rate
0.05	0.54
0.10	0.36
0.25	0.16
0.50	0.06
0.75	0.02
0.90	0.006
0.95	0.003

$$\alpha = 0.05 \qquad \beta = 0.80$$

Exemplo: Correlação

- Suponha que estamos interessados em medir a associação entre duas variáveis contínuas, e para isso vamos selecionar três amostras com diferentes tamanhos.

- N = {1.000, 10.000, 100.000}

- Suponha também que conhecemos a lei de probabilidade que essas duas variáveis seguem.

Exemplo: Correlação

- Elas segue a seguinte normal multivariada:

$$\begin{pmatrix} X \\ Y \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0,01 \\ 0,01 & 1 \end{pmatrix}\right)$$

- Logo

$$\mathbb{C}orr(X,Y) = 0.01$$

Exemplo: Correlação

```
# Gerando as amostras

require(mvtnorm)
set.seed(123456) # Fixando a semente

X1 <- rmvnorm(n=1000, sigma = matrix(c(1,0.01,0.01,1),2,2))
X2 <- rmvnorm(n=10000, sigma = matrix(c(1,0.01,0.01,1),2,2))
X3 <- rmvnorm(n=100000, sigma = matrix(c(1,0.01,0.01,1),2,2))
```

Exemplo: Correlação

```
## N = 1,000

cor.test(x=X1[,1], y=X1[,2])$p.value
## 0.03627

## N = 10,000

cor.test(x=X2[,1], y=X2[,2])$p.value
## 0.278

## N = 100,000

cor.test(x=X3[,1], y=X3[,2])$p.value
## 0.0001761
```

$$\begin{cases} H_0: & \rho = 0 \\ H_A: & \rho \neq 0 \end{cases}$$

O problema já vem sendo discutido:

- Ioannidis (2005, Plos Medicine):

  “Why Most Published Research Findings Are False”

- Siegfried, T. (2010, ScienceNews)

  “Odds Are, It's Wrong: Science fails to face the shortcomings of statistics,”

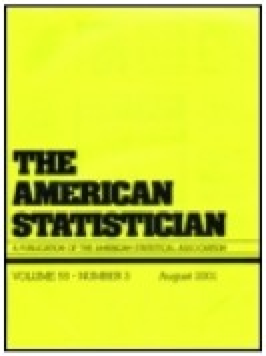
- Nuzzo (2014, Nature)

  “Scientific method: Statistical errors”

- Baker (2016, Nature)

  “Statisticians issue warning over misuse of P values”

Posição da Associação Americana de Estatística (ASA) sobre o valor-p



The American Statistician

ISSN: 0003-1305 (Print) 1537-2731 (Online) Journal homepage: http://amstat.tandfonline.com/loi/utas20

The ASA's statement on p-values: context, process, and purpose

Ronald L. Wasserstein & Nicole A. Lazar

To cite this article: Ronald L. Wasserstein & Nicole A. Lazar (2016): The ASA's statement on p-values: context, process, and purpose, The American Statistician, DOI: 10.1080/00031305.2016.1154108

To link to this article: http://dx.doi.org/10.1080/00031305.2016.1154108

Princípios propostos pela ASA

1. Valor-p pode indicar o quão incompatíveis são os dados de um modelo pré especificado;
2. Valor-p não mede a probabilidade da hipotese nula ser verdadeira, ou a probabilidade de que os dados foram gerados totalmente ao acaso;
3. Conclusões científicas e decisões políticas ou de negócios NÃO devem ser baseadas somente em se o valor-p passa ou não um valor pré-especificado;
4. Inferência requer uma descrição completa e transparente;
5. Um valor-p, ou significância estatística, NÃO medem o tamanho de um efeito ou importância de um resultado;
6. Sozinho, o valor-p NÃO fornece uma boa medida de evidência sobre um modelo ou hipótese;

Alternativas?

- Reportar incertezas (estimativas e seus intervalos)
- False-discovery rate
- Model approach
  - Fator de Bayes
  - medidas de comparação de modelos {LRT, AIC, BIC, DIC, WAIC, (bla bla bla)-IC, etc}
  - Valor-e (*e-value*)
  - FBST {Carlinhos Pereira, USP}
- Etc.

Por que elas não colaram?

- Benjamin e Berger argumentam que nenhuma alternativa ao valor-p conseguiu apoio geral por uma das duas razões:

1) A alternativa é mais complicada que o valor-p

2) A alternativa não é aceita por ambas escolas de pensamento estatístico: bayesianos e frequentistas

Voltando ao exemplo da correlação

- Lembrando que as duas variáveis seguem a seguinte lei de probabilidades:

$$\begin{pmatrix} X \\ Y \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0,01 \\ 0,01 & 1 \end{pmatrix}\right)$$

- Logo

$$\mathbb{C}orr(X,Y) = 0.01$$

Exemplo: Correlação

```
# Gerando as amostras

require(mvtnorm)
set.seed(123456) # Fixando a semente

X1 <- rmvnorm(n=1000, sigma = matrix(c(1,0.01,0.01,1),2,2))
X2 <- rmvnorm(n=10000, sigma = matrix(c(1,0.01,0.01,1),2,2))
X3 <- rmvnorm(n=100000, sigma = matrix(c(1,0.01,0.01,1),2,2))
```

Exemplo: Correlação

```
## N = 1,000

cor.test(x=X1[,1], y=X1[,2])

##
##  Pearson's product-moment correlation
##
## data:  X1[, 1] and X1[, 2]
## t = 2.0967, df = 998, p-value = 0.03627
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.004248519 0.127693006
## sample estimates:
##       cor
## 0.06622416
```

$$\begin{cases} H_0: & \rho = 0 \\ H_A: & \rho \neq 0 \end{cases}$$

Exemplo: Correlação

```
## N = 10,000

cor.test(x=X2[,1], y=X2[,2])

##
##  Pearson's product-moment correlation
##
## data:  X2[, 1] and X2[, 2]
## t = -1.0848, df = 9998, p-value = 0.278
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  -0.030442169  0.008753359
## sample estimates:
##         cor
## -0.01084857
```

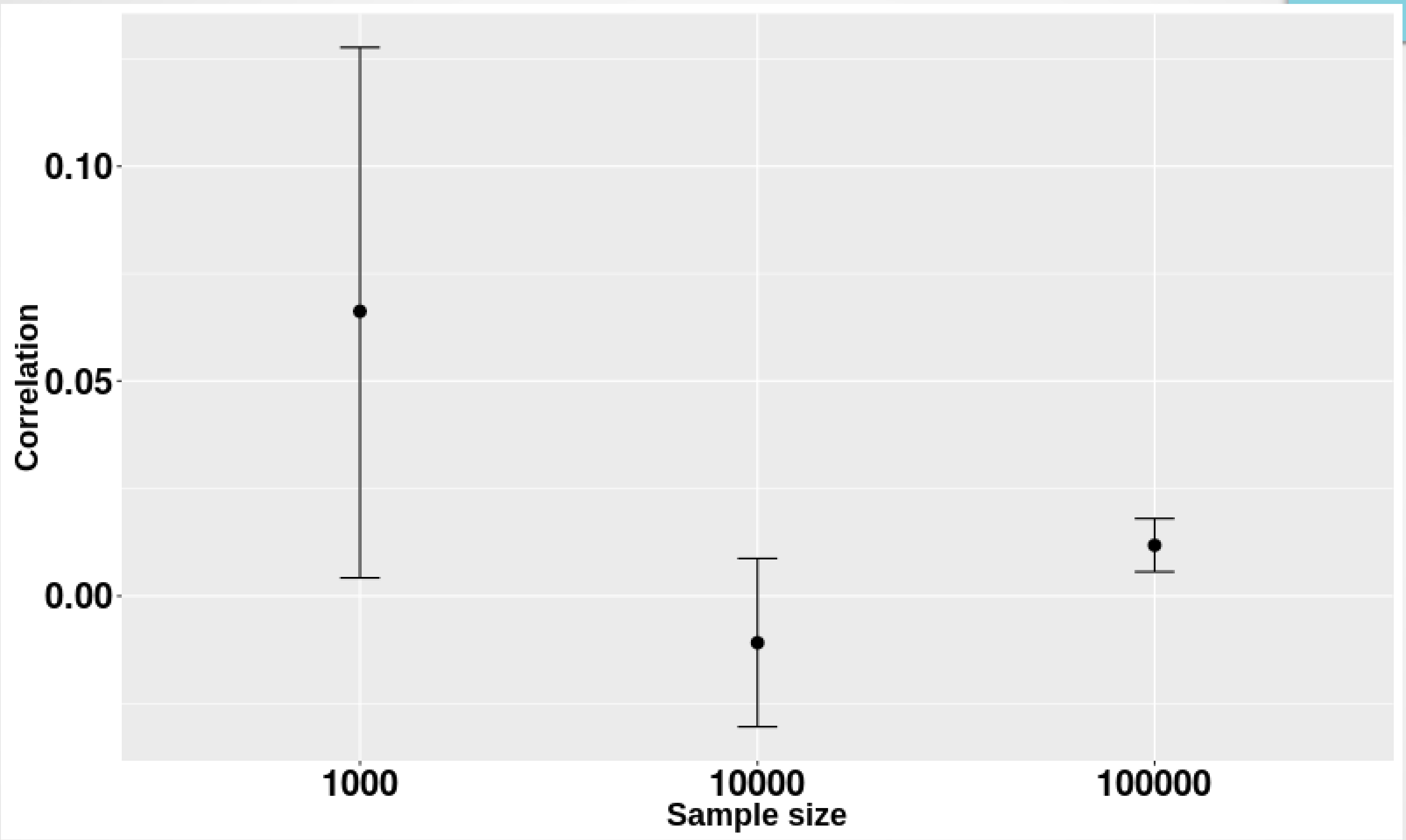
$$\begin{cases} H_0: & \rho = 0 \\ H_A: & \rho \neq 0 \end{cases}$$

Exemplo: Correlação

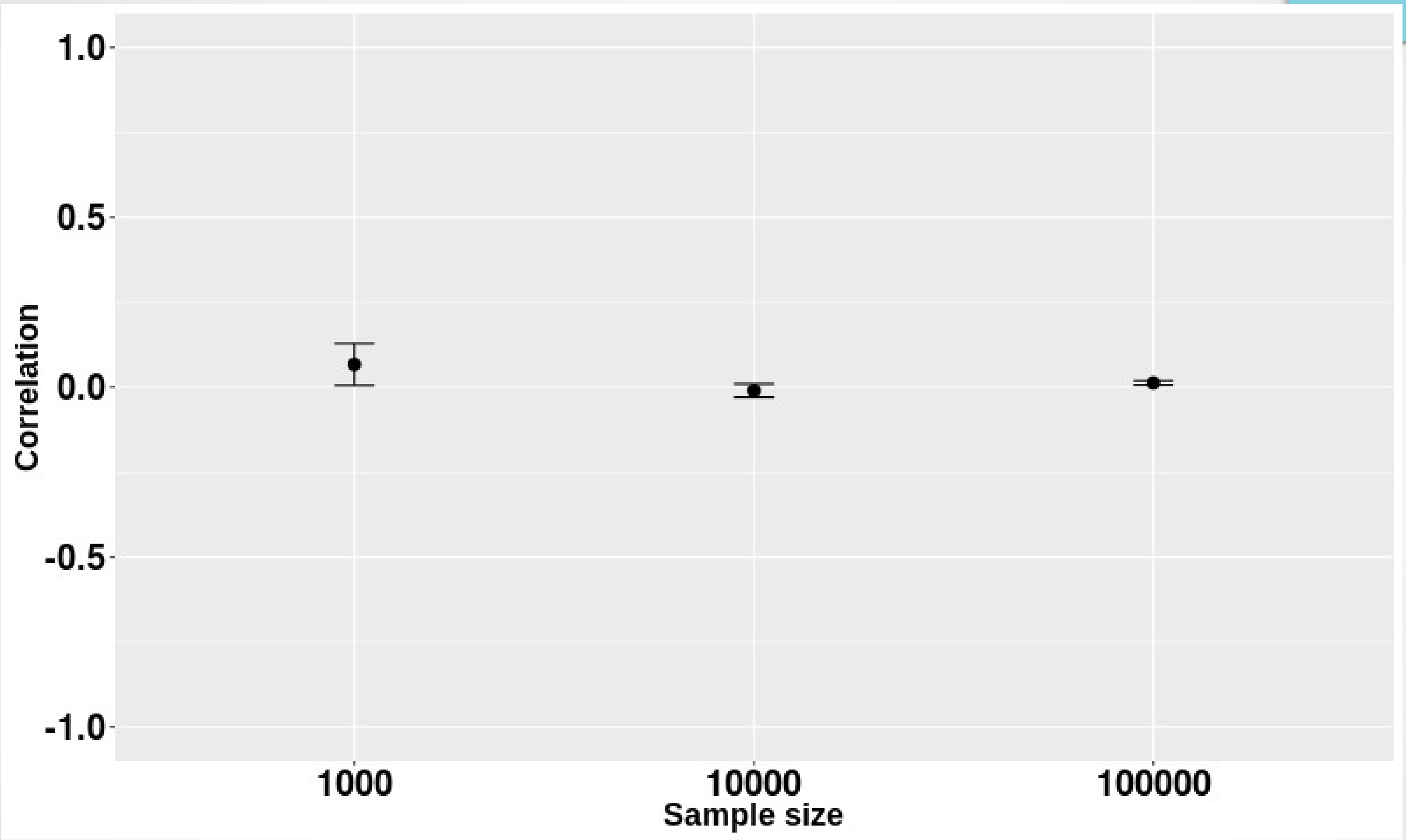
```
## N = 100,000

cor.test(x=X3[,1], y=X3[,2])

##
##  Pearson's product-moment correlation
##
## data:  X3[, 1] and X3[, 2]
## t = 3.7511, df = 99998, p-value = 0.0001761
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  0.005663867 0.018058051
## sample estimates:
##       cor
## 0.01186141
```

$$\begin{cases} H_0: & \rho = 0 \\ H_A: & \rho \neq 0 \end{cases}$$

Exemplo: Correlação



Exemplo: Correlação



Por que ensinamos $P < 0.05$?

Q: Why do so many colleges and grad schools teach $p = .05$?

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.

“The p-value was never intended to be a substitute for scientific reasoning”

Ron Wassertein, Executive director, ASA

Obrigado pela atenção!

Leo Bastos
PROCC/Fiocruz
Leonardo.bastos@fiocruz.br