

# Significância e Poder

Testes de Hipóteses, cálculo amostral e p-valor

Felipe Figueiredo

- 1 Testes de Hipóteses
  - Hipóteses
  - Poder estatístico
  - Significância Estatística
  - O p-valor é...
  - Exercício
  - O p-valor não é...
- 2 Aprofundamento
  - Aprofundamento

## Discussão da leitura obrigatória da aula passada

## 1 Testes de Hipóteses

- Hipóteses
- Poder estatístico
- Significância Estatística
- O p-valor é...
- Exercício
- O p-valor não é...

## 2 Aprofundamento

- Aprofundamento

## Livro texto - Parte III - Introduction to p values

*"I've put it off for nine chapters, but I can't delay any longer.  
**It's time to confront P values.** (...)*

*If you've had any exposure to statistics before, you've probably already heard  
about P values and statistical significance.  
It's time to learn what these phrases really mean. (...)*

*These chapters explain P values generally, without explaining any particular  
statistical tests in any detail"*

Motulsky, 1995  
(grifos e quebras meus)

## Abertura de “A Divina Comédia”

*“A meio do caminho, ou seja, da duração expectável de sua vida, Dante, consciente de se haver desviado do reto procedimento, encontra-se perdido numa alegórica ‘Selva Perdida’.*

*Encontra aí a figura de Virgílio, o poeta latino que (...) vem se lhe oferecer como guia para o Inferno e o Purgatório onde, pelo exemplo dos pecadores e de suas penas, Dante poderá encontrar o caminho da sua salvação.”*

Dante Alighieri, 1320

Sempre que mensuramos algo... observamos diferenças

Estas são **devidas ao efeito estudado?**

Ou são causadas pela variabilidade,  
erros metrológicos, etc?

## Exemplo 1

### Exemplo 1

Um neurologista está testando o efeito de uma droga no tempo de resposta de um certo estímulo neurológico. Para isto, ele injeta uma dose da droga em **100** ratos, cria os estímulos neurológicos e observa o tempo de resposta em cada animal.

O neurologista sabe que o tempo de resposta médio de ratos que não receberam a droga é de **1.2 segundos**.

O tempo de resposta médio dos ratos injetados foi de **1.05 segundos**, com desvio padrão amostral de **0.5 segundos**.

**Você acha que a droga tem efeito no tempo de resposta do estímulo?**

Fonte: Khan Academy

Significância

Felipe  
Figueiredo

Testes de  
Hipóteses

Hipóteses

Poder

Significância

O p-valor é...

Exercício

O p-valor não é...

Aprofundamento



## Exemplo 1

### Exemplo 1

Um neurologista está testando o efeito de uma droga no tempo de resposta de um certo estímulo neurológico. Para isto, ele injeta uma dose da droga em **100** ratos, cria os estímulos neurológicos e observa o tempo de resposta em cada animal.

O neurologista sabe que o tempo de resposta médio de ratos que não receberam a droga é de **1.2 segundos**.

O tempo de resposta médio dos ratos injetados foi de **1.05 segundos**, com desvio padrão amostral de **0.5 segundos**.

Você tem informações suficientes para construir um IC em torno de  $\bar{x}$ ?

Fonte: Khan Academy

Significância

Felipe  
Figueiredo

Testes de  
Hipóteses

Hipóteses

Poder

Significância

O p-valor é...

Exercício

O p-valor não é...

Aprofundamento

## Exemplo 1

Um neurologista está testando o efeito de uma droga no tempo de resposta de um certo estímulo neurológico.

Tempo de resposta típico = 1.2 s.

$$n = 100$$

$$\bar{x} = 1.05 \text{ s}$$

$$s = 0.5 \text{ s}$$

## IC da média

$$\text{IC} = 0.9508 \text{ até } 1.1492$$

## Exemplo 1

Um neurologista está testando o efeito de uma droga no tempo de resposta de um certo estímulo neurológico.

Tempo de resposta típico = 1.2 s.

$$n = 100$$

$$\bar{x} = 1.05 \text{ s}$$

$$s = 0.5 \text{ s}$$

## IC da média

IC = 0.9508 até 1.1492

## IC da diferença entre $\bar{x}$ e 1.2

$\bar{d} = -0.1500$ , IC = -0.2492 até -0.0508

## Exemplo 1

Um neurologista está testando o efeito de uma droga no tempo de resposta de um certo estímulo neurológico.

Tempo de resposta típico = 1.2 s.

$$n = 100$$

$$\bar{x} = 1.05 \text{ s}$$

$$s = 0.5 \text{ s}$$

## Teste de significância entre $\bar{x}$ e 1.2

$$p = 0.0034$$

É hoje!

## Significância

Felipe  
Figueiredo

## Testes de Hipóteses

### Hipóteses

Poder

Significância

O p-valor é...

Exercício

O p-valor não é...

## Aprofundamento



*Laaaaadies aaaaand gentlemen...*

- Podemos tomar decisões baseado nos dados de um experimento (amostra).
- Para isto, precisamos de um critério sistemático e rigoroso que possa aferir o quanto os dados suportam esta decisão.
- Usando os conceitos de probabilidades, poderemos ainda calcular a probabilidade de que esta decisão esteja errada.

**Hipóteses devem ser falseáveis, portanto formuladas como afirmações.**

Hipóteses estatísticas podem ser verdadeiras ou falsas...  
... não existe “mais ou menos” significativo!

Não existe “mais ou menos” significativo...

<u>P-VALUE</u>	<u>INTERPRETATION</u>
0.001	HIGHLY SIGNIFICANT
0.01	
0.02	
0.03	
0.04	SIGNIFICANT
0.049	
0.050	
0.051	OH CRAP. REDO CALCULATIONS.
0.06	
0.07	ON THE EDGE OF SIGNIFICANCE
0.08	
0.09	
0.099	HIGHLY SUGGESTIVE, SIGNIFICANT AT THE P<0.10 LEVEL
≥0.1	
	HEY, LOOK AT THIS INTERESTING SUBGROUP ANALYSIS

Fonte: <https://xkcd.com/1478/>

## Significância

Felipe  
Figueiredo

## Testes de Hipóteses

Hipóteses

Poder

Significância

O p-valor é...

Exercício

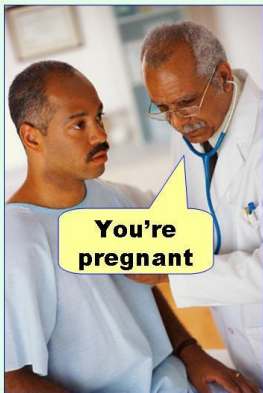
O p-valor não é...

## Aprofundamento



... assim como não existe “mais ou menos” grávido!

**Type I error**  
(false positive)



**Type II error**  
(false negative)



Significância

Felipe  
Figueiredo

Testes de  
Hipóteses

Hipóteses

Poder

Significância

O p-valor é...

Exercício

O p-valor não é...

Aprofundamento

## Definição

Em Estatística, uma **hipótese** é uma afirmação sobre uma característica de uma população, tipicamente o valor de um parâmetro.

## Definição

Um **teste de hipóteses** (ou teste de significância) é um procedimento sistemático para testar uma afirmação sobre uma característica de uma população.

### Exemplo 1

Um neurologista está testando o efeito de uma droga no tempo de resposta de um certo estímulo neurológico. Para isto, ele injeta uma dose da droga em 100 ratos, cria os estímulos neurológicos e observa o tempo de resposta em cada animal.

O neurologista sabe que o tempo de resposta médio de ratos que não receberam a droga é de 1.2 segundos.

O tempo de resposta médio dos ratos injetados foi de 1.05 segundos, com desvio padrão amostral de 0.5 segundos.

Você acha que a droga tem efeito no tempo de resposta do estímulo?

Fonte: Khan Academy

## Exemplo 1

Um neurologista está testando o efeito de uma droga no tempo de resposta de um certo estímulo neurológico.

Tempo de resposta típico = 1.2 s.

$$n = 100$$

$$\bar{x} = 1.05 \text{ s}$$

$$s = 0.5 \text{ s}$$

## Pense...

Que possíveis conclusões você pode chegar com esse experimento?

## Exemplo 1

Um neurologista está testando o efeito de uma droga no tempo de resposta de um certo estímulo neurológico.

Tempo de resposta típico = 1.2 s.

$$n = 100$$

$$\bar{x} = 1.05 \text{ s}$$

$$s = 0.5 \text{ s}$$

## Pense...

Como você formularia a hipótese do exemplo anterior?

Uma hipótese estatística deve ser testável frente a dados obtidos de um experimento.

### Pergunta

A droga tem efeito no tempo de resposta do estímulo?

### Hipótese

A droga influencia (altera/afeta) o tempo de resposta do estímulo.

### 1 teste = 2 hipóteses

Um teste de hipóteses envolve a formulação de uma *hipótese nula* e uma *hipótese alternativa*.

- A hipótese nula ( $H_0$ ) é a hipótese que não há efeito real.
- A hipótese alternativa ( $H_1$  ou  $H_a$ ) é a de que há efeito real **que pode ser detectado** com o experimento.
  - Obs: (em geral) hipótese de interesse científico

### *Danger Will Robinson...*

A lógica do teste de hipóteses é o **inverso** do que se esperaria intuitivamente.

Isto é, ao invés de testar a hipótese de interesse, vamos *testar a hipótese nula* – e tentar rejeitá-la.

**Mantenha isso em mente daqui a para a frente.**



### Roteiro

- 1 Identificar a afirmação a ser testada e expressá-la em forma simbólica
- 2 Expressar em forma simbólica a afirmação que deve ser verdadeira, caso a afirmação de interesse seja falsa

## Exemplo 1

Um neurologista está testando o efeito de uma droga no tempo de resposta de um certo estímulo neurológico.

Tempo de resposta típico = 1.2 s.

$$n = 100$$

$$\bar{x} = 1.05 \text{ s}$$

$$s = 0.5 \text{ s}$$

Significância

Felipe  
Figueiredo

Testes de  
Hipóteses

Hipóteses

Poder

Significância

O p-valor é...

Exercício

O p-valor não é...

Aprofundamento

# Quais são as variáveis?

## Exemplo 1

Um neurologista está testando o efeito de uma droga no tempo de resposta de um certo estímulo neurológico.

Tempo de resposta típico = 1.2 s.

$n = 100$

$\bar{x} = 1.05$  s

$s = 0.5$  s

## Modelo

variável dependente  $\sim$  variável independente

tempo de resposta  $\sim$  ratos (grupo)

Significância

Felipe  
Figueiredo

Testes de  
Hipóteses

Hipóteses

Poder

Significância

O p-valor é...

Exercício

O p-valor não é...

Aprofundamento

## Quais são as hipóteses estatísticas?

### Exemplo 1

Um neurologista está testando o efeito de uma droga no tempo de resposta de um certo estímulo neurológico.

Tempo de resposta típico = 1.2 s.

$$n = 100$$

$$\bar{x} = 1.05 \text{ s}$$

$$s = 0.5 \text{ s}$$

Significância

Felipe  
Figueiredo

Testes de  
Hipóteses

Hipóteses

Poder

Significância

O p-valor é...

Exercício

O p-valor não é...

Aprofundamento

# Quais são as hipóteses estatísticas?

## Exemplo 1

Um neurologista está testando o efeito de uma droga no tempo de resposta de um certo estímulo neurológico.

Tempo de resposta típico = 1.2 s.

$$n = 100$$

$$\bar{x} = 1.05 \text{ s}$$

$$s = 0.5 \text{ s}$$

## Hipóteses

$$H_0 : \mu = 1.2$$

$$H_1 : \mu \neq 1.2$$

Significância

Felipe  
Figueiredo

Testes de  
Hipóteses

Hipóteses

Poder

Significância

O p-valor é...

Exercício

O p-valor não é...

Aprofundamento

Em geral...

- Se  $H_1$  é do tipo  $\neq$ , o teste é bicaudal (ou bilateral).
- Se  $H_1$  é do tipo  $<$ , o teste é unicaudal (ou unilateral) à esquerda.
- Se  $H_1$  é do tipo  $>$ , o teste é unicaudal à direita.

## 1 Testes de Hipóteses

- Hipóteses
- **Poder estatístico**
- Significância Estatística
- O p-valor é...
- Exercício
- O p-valor não é...

## 2 Aprofundamento

- Aprofundamento

## Queremos

- planejar estudos com poder suficiente para detectar o efeito investigado
- para isso é necessário especificar<sup>1</sup>
  - o tamanho do efeito desejado
  - quanto poder é “suficiente”
  - qual é a significância aceitável

---

<sup>1</sup> com essas especificações é possível calcular o tamanho do estudo (“qual é o N?”) ◀ ≡ ▶ ≡ 🔍 ↺



Vamos entender os conceitos envolvidos...

... por uma analogia.

Quanto menor o  $N$ , mais difícil detectar o efeito

- Sua visão tem um poder de detecção predefinido
- **P: Quantos objetos você consegue contar na imagem abaixo?**

Mais tempo, mais objetos detectados?

- ... em 15 segundos?
- ... em 60 segundos?

Significância

Felipe  
Figueiredo

Testes de  
Hipóteses

Hipóteses

Poder

Significância

O p-valor é...

Exercício

O p-valor não é...

Aprofundamento

Quanto menor o  $N$ , mais difícil detectar o efeito

Significância

Felipe  
Figueiredo

Testes de  
Hipóteses

Hipóteses

Poder

Significância

O p-valor é...

Exercício

O p-valor não é...

Aprofundamento

- Sua visão tem um poder de detecção predefinido
- **P: Quantos objetos você consegue contar na imagem abaixo?**

Mais tempo, mais objetos detectados?



- ... em 15 segundos?
- ... em 60 segundos?

Quanto menor o  $N$ , mais difícil detectar o efeito

- Sua visão tem um poder de detecção predefinido
- **P: Quantos objetos você consegue contar na imagem abaixo?**

Mais tempo, mais objetos detectados?

- ... em 15 segundos?
- ... em 60 segundos?

Significância

Felipe  
Figueiredo

Testes de  
Hipóteses

Hipóteses

Poder

Significância

O p-valor é...

Exercício

O p-valor não é...

Aprofundamento

Quanto menor o  $N$ , mais difícil detectar o efeito

Significância

Felipe  
Figueiredo

Testes de  
Hipóteses

Hipóteses

Poder

Significância

O p-valor é...

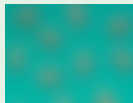
Exercício

O p-valor não é...

Aprofundamento

- Sua visão tem um poder de detecção predefinido
- **P: Quantos objetos você consegue contar na imagem abaixo?**

Mais tempo, mais objetos detectados?



- ... em 15 segundos?
- ... em 60 segundos?

Qual é o impacto do N no poder do estudo?

Quanto menor o efeito, mais difícil detectá-lo

- Sua visão tem um poder de detecção predefinido
- **P: Quantos objetos você consegue contar na imagem abaixo?**

Efeito pequeno



Efeito grande



Maior tamanho de efeito, mais fácil contar os objetos?

Significância

Felipe  
Figueiredo

Testes de  
Hipóteses

Hipóteses

Poder

Significância

O p-valor é...

Exercício

O p-valor não é...

Aprofundamento

Se você precisa detectar um efeito pequeno, precisa um N:

- 1 maior?
- 2 menor?
- 3 indiferente?



Quanto mais variabilidade, mais difícil detectar o efeito

- Sua visão tem um poder de detecção predefinido
- Quantos objetos você consegue contar na imagem abaixo?

Muita variabilidade



Menos variabilidade



Menos variabilidade, mais fácil contar os objetos?

Significância

Felipe  
Figueiredo

Testes de  
Hipóteses

Hipóteses

Poder

Significância

O p-valor é...

Exercício

O p-valor não é...

Aprofundamento

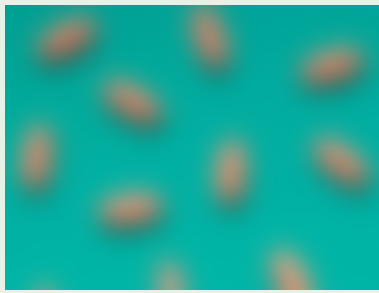
Quanto mais variabilidade, mais difícil detectar o efeito

- Sua visão tem um poder de detecção predefinido
- Quantos objetos você consegue contar na imagem abaixo?

Muita variabilidade



Pouca variabilidade



E agora?

Significância

Felipe  
Figueiredo

Testes de  
Hipóteses

Hipóteses

Poder

Significância

O p-valor é...

Exercício

O p-valor não é...

Aprofundamento

Qual é o impacto da variabilidade no poder do estudo?

Sempre que mensuramos algo... observamos diferenças

Estas são **devidas ao efeito estudado?**

Ou são causadas pela variabilidade,  
erros metrológicos, etc?

- O mesmo ocorre com os métodos estatísticos (ferramenta)

- O mesmo ocorre com os métodos estatísticos (ferramenta)
- Você a usa para investigar a existência de um efeito

- O mesmo ocorre com os métodos estatísticos (ferramenta)
- Você a usa para investigar a existência de um efeito
- A capacidade da ferramenta em detectá-lo depende...
  - da variabilidade presente
  - do tamanho do efeito relevante
  - do tamanho do estudo

- O mesmo ocorre com os métodos estatísticos (ferramenta)
- Você a usa para investigar a existência de um efeito
- A capacidade da ferramenta em detectá-lo depende...
  - da variabilidade presente ( $s$ ) (aula 3)
  - do tamanho do efeito relevante ( $d$ ) (aula 5)
  - do tamanho do estudo ( $N$ )



## Análise de poder / cálculo amostral

Para que o estudo seja capaz de detectar o efeito desejado

- Estipulamos um N e calculamos o poder
- Estipulamos o poder e calculamos o N

## Poder para cada tamanho de efeito

**Table 12.2.** A Power Analysis of Example 12.1

Relative Risk	Power
0.95	11%
0.90	30%
0.85	60%
0.80	84%
0.75	97%

## Significância

Felipe  
Figueiredo

## Testes de Hipóteses

Hipóteses

Poder

Significância

O p-valor é...

Exercício

O p-valor não é...

## Aprofundamento

## Diferença em...

### Desfecho qualitativo

- **Proporção na população**  
ex.  $P = 80\%$
- **Tamanho do efeito relevante**  
diminuir de 80% para 70%

$d = 10\%$

### Desfecho quantitativo

- **Média da população**  
 $\mu = 1.2$  segundos
- **Variabilidade da população**  
 $\sigma = 0.3$  segundos
- **Tamanho do efeito relevante**  
diminuir de 1.2 para 1.1 segundos

$d = 0.1$  segundos

## Significância

Felipe  
Figueiredo

## Testes de Hipóteses

Hipóteses

Poder

Significância

O p-valor é...

Exercício

O p-valor não é...

## Aprofundamento

- 1 Testes de Hipóteses
  - Hipóteses
  - Poder estatístico
  - **Significância Estatística**
  - O p-valor é...
  - Exercício
  - O p-valor não é...
- 2 Aprofundamento
  - Aprofundamento

- A **significância** do estudo deve ser arbitrada antes do experimento (planejamento)
- Está associada aos erros induzidos pela variabilidade experimental
- Ou seja, mesmo fazendo tudo certo, você pode ser induzido a chegar numa conclusão errada ao acaso!
- Isso pode ocorrer de duas maneiras diferentes...

Cada tipo de erro pode ter um “custo” diferente

- Dizer que um paciente não está infectado, quando ele está.
- Dizer que um paciente está infectado, quando ele não está.

O custo de cada tipo de possível erro só pode ser avaliado caso a caso.

Significância

Felipe  
Figueiredo

Testes de  
Hipóteses

Hipóteses

Poder

Significância

O p-valor é...

Exercício

O p-valor não é...

Aprofundamento

## Definição

Um **erro do tipo I** ocorre se a hipótese nula for rejeitada quando é verdadeira.

## Definição

Um **erro do tipo II** ocorre se a hipótese não for rejeitada quando for falsa.

## A questão importante aqui é:

MESMO SE a hipótese nula **for verdadeira**, ainda assim você pode observar (ao acaso) uma diferença como resultado do experimento.

(ex., muita variabilidade, amostras pequenas, etc.).

**Isso** é o erro tipo I. Trabalhamos para que isso seja raro (não mais que 5% das vezes).

# Tipos de erros em testes de hipóteses

**Table 15.1.** False Negatives and Positives in Diagnostic Tests

Diagnostic Test	Disease Is Really Present	Disease Is Really Absent
Test positive	No error (true positive)	False positive
Test negative	False negative	No error (true negative)

## Lab

Se o paciente realmente tem a doença, qual é a chance do teste diagnóstico retornar corretamente um resultado positivo? **A resposta é a sensibilidade.** Se o teste pode detectar várias doenças, a sensibilidade depende de **qual doença** está sendo investigada.

## Significância

Felipe  
Figueiredo

## Testes de Hipóteses

Hipóteses

Poder

Significância

O p-valor é...

Exercício

O p-valor não é...

## Aprofundamento



# Tipos de erros em testes de hipóteses

**Table 15.1.** False Negatives and Positives in Diagnostic Tests

Diagnostic Test	Disease Is Really Present	Disease Is Really Absent
Test positive	No error (true positive)	False positive
Test negative	False negative	No error (true negative)

## Lab

Se o paciente realmente tem a doença, qual é a chance do teste diagnóstico retornar corretamente um resultado positivo? **A resposta é a sensibilidade.** Se o teste pode detectar várias doenças, a sensibilidade depende de **qual doença** está sendo investigada.

## Stats

Se realmente existe uma diferença ( $\Delta$ ) entre as médias das populações qual é a chance que a análise de uma amostragem aleatória ser capaz de diferença estatística? **A resposta é o poder.** O poder depende do **tamanho do efeito** ( $\Delta$ ) presumido.

**Table 15.2.** Type I and Type II Errors in Statistical Tests

Statistical Test	Populations Have Different Means (or Proportions)	Populations Have Identical Means (or Proportions)
Significant difference	No error	Type I error
No significant difference	Type II error	No error

## Significância

Felipe  
Figueiredo

## Testes de Hipóteses

Hipóteses

Poder

Significância

O p-valor é...

Exercício

O p-valor não é...

## Aprofundamento

## Definição

O **nível de significância** de um teste de hipótese é sua probabilidade máxima admissível para cometer um erro do tipo I. Ele é denotado por  $\alpha$ .

$$\text{confiança} = 1 - \alpha$$

## Definição

A probabilidade de se cometer um erro do tipo II é denotada por  $\beta$ .

$$\text{poder} = 1 - \beta$$

São necessários para um teste de hipóteses:

- As hipóteses nula e alternativa
- O nível de significância
- A região crítica (tipo de teste)
- A estatística de teste (softwares especializados)

## Observação

O teste unicaudal **divide** a probabilidade de erro à esquerda (valores menores) e à direita (valores maiores).

Assim, 5% de significância num teste unicaudal corresponde à 2.5% (metade) da significância bicaudal.

Mais detalhes no cap 10.

Significância

Felipe  
Figueiredo

Testes de  
Hipóteses

Hipóteses

Poder

Significância

O p-valor é...

Exercício

O p-valor não é...

Aprofundamento

### Importante

Observe que o teste de hipótese nunca deve **aceitar** uma hipótese nula, apenas rejeitá-la ou deixar de rejeitá-la.

## 1 Testes de Hipóteses

- Hipóteses
- Poder estatístico
- Significância Estatística
- O p-valor é...
- Exercício
- O p-valor não é...

## 2 Aprofundamento

- Aprofundamento

## Definição

Assumindo que a hipótese nula seja verdadeira, o **p-valor** de um teste de hipóteses é a probabilidade de se obter uma estatística amostral com valores **tão extremos, ou mais extremos** que aquele observado.

O p-valor **é**:

- A probabilidade (condicional) de se observar o resultado ao acaso **dado que** a  $H_0$  é verdadeira.
- Uma medida da força da evidência **contra** a  $H_0$ .

## Como utilizar

- Quanto menor o p-valor, mais evidências para rejeitar a hipótese nula.
- O ponto de corte mais utilizado é a significância de 5%
- Assim, qualquer  $p \leq 0.05$  é estatisticamente significativo.

## Exemplo 1

### Exemplo 1

Um neurologista está testando o efeito de uma droga no tempo de resposta de um certo estímulo neurológico. Para isto, ele injeta uma dose da droga em **100** ratos, cria os estímulos neurológicos e observa o tempo de resposta em cada animal.

O neurologista sabe que o tempo de resposta médio de ratos que não receberam a droga é de **1.2 segundos**.

O tempo de resposta médio dos ratos injetados foi de **1.05 segundos**, com desvio padrão amostral de **0.5 segundos**.

**Você acha que a droga tem efeito no tempo de resposta do estímulo?**

Fonte: Khan Academy

Significância

Felipe  
Figueiredo

Testes de  
Hipóteses

Hipóteses

Poder

Significância

O p-valor é...

Exercício

O p-valor não é...

Aprofundamento



## Pense...

- A hipótese científica é que a droga afeta o tempo de resposta.
- Como você formularia a hipótese estatística ( $H_1$ )?
  - 1  $H_0 : \mu = 1.2, H_1 : \mu \geq 1.2$  (teste unicaudal à direita)
  - 2  $H_0 : \mu = 1.2, H_1 : \mu < 1.2$  (teste unicaudal à esquerda)
  - 3  $H_0 : \mu = 1.2, H_1 : \mu \neq 1.2$  (teste bicaudal)
  - 4  $H_0 : \mu \geq 1.2, H_1 : \mu = 1.2$  (teste unicaudal à esquerda)

Resposta:

## Exemplo 1

Significância

Felipe  
Figueiredo

Testes de  
Hipóteses

Hipóteses

Poder

Significância

O p-valor é...

Exercício

O p-valor não é...

Aprofundamento

### Pense...

- A hipótese científica é que a droga afeta o tempo de resposta.
- Como você formularia a hipótese estatística ( $H_1$ )?
  - 1  $H_0 : \mu = 1.2, H_1 : \mu \geq 1.2$  (teste unicaudal à direita)
  - 2  $H_0 : \mu = 1.2, H_1 : \mu < 1.2$  (teste unicaudal à esquerda)
  - 3  $H_0 : \mu = 1.2, H_1 : \mu \neq 1.2$  (teste bicaudal)
  - 4  $H_0 : \mu \geq 1.2, H_1 : \mu = 1.2$  (teste unicaudal à esquerda)

Resposta: **Opção 3**

## Exemplo 1

- Dados:  $\mu = 1.2, \bar{x} = 1.05, s = 0.5, n = 100$
- $H_0 : \mu = 1.2, H_1 : \mu \neq 1.2$  (teste bicaudal)
  - O teste Z retorna  $p = 0.0027^2$
  - O teste t retorna  $p = 0.0034^3$

### Significância

Felipe  
Figueiredo

### Testes de Hipóteses

Hipóteses

Poder

Significância

O p-valor é...

Exercício

O p-valor não é...

### Aprofundamento

<sup>2</sup> Premissas fortes: Normal, N grande,  $\sigma$  conhecido, etc.

<sup>3</sup> Usado em geral, menos premissas

## Exemplo 1

- Dados:  $\mu = 1.2$ ,  $\bar{x} = 1.05$ ,  $s = 0.5$ ,  $n = 100$
- $H_0 : \mu = 1.2$ ,  $H_1 : \mu \neq 1.2$  (teste bicaudal)
  - O teste Z retorna  $p = 0.0027^2$
  - O teste t retorna  $p = 0.0034^3$
- Como  $p < 0.05$ , há evidências para rejeitar  $H_0$ .

<sup>2</sup> Premissas fortes: Normal, N grande,  $\sigma$  conhecido, etc.

<sup>3</sup> Usado em geral, menos premissas

### Significância

Felipe  
Figueiredo

### Testes de Hipóteses

Hipóteses

Poder

Significância

O p-valor é...

Exercício

O p-valor não é...

### Aprofundamento

### Exemplo 1

- Dados:  $\mu = 1.2, \bar{x} = 1.05, s = 0.5, n = 100$
- $H_0 : \mu = 1.2, H_1 : \mu \neq 1.2$  (teste bicaudal)
  - O teste Z retorna  $p = 0.0027^2$
  - O teste t retorna  $p = 0.0034^3$
- Como  $p < 0.05$ , há evidências para rejeitar  $H_0$ .

## Resultado

O tempo de resposta médio é **significativamente** diferente de 1.2 s ( $p = 0.0034$ ).

<sup>2</sup> Premissas fortes: Normal, N grande,  $\sigma$  conhecido, etc.

<sup>3</sup> Usado em geral, menos premissas

### Exemplo 1

- Dados:  $\mu = 1.2, \bar{x} = 1.05, s = 0.5, n = 100$
- $H_0 : \mu = 1.2, H_1 : \mu \neq 1.2$  (teste bicaudal)
  - O teste Z retorna  $p = 0.0027^2$
  - O teste t retorna  $p = 0.0034^3$
- Como  $p < 0.05$ , há evidências para rejeitar  $H_0$ .

## Resultado

O tempo de resposta médio é **significativamente** diferente de 1.2 s ( $p = 0.0034$ ).

## Conclusão

(...) há evidências que a droga altera o tempo (...) de resposta (...).

<sup>2</sup> Premissas fortes: Normal, N grande,  $\sigma$  conhecido, etc.

<sup>3</sup> Usado em geral, menos premissas

Sempre que mensuramos algo... observamos diferenças

Estas são **devidas ao efeito estudado?**

Ou são causadas pela variabilidade,  
erros metrológicos, etc?

## 1 Testes de Hipóteses

- Hipóteses
- Poder estatístico
- Significância Estatística
- O p-valor é...
- **Exercício**
- O p-valor não é...

## 2 Aprofundamento

- Aprofundamento



## Exemplo 2

Uma indústria farmacêutica especifica que em certo analgésico a quantidade média de ácido acetil salicílico deve ser 5.5 gramas por comprimido. A indústria suspeita que houve problemas na produção de um determinado lote e que, nesse lote, a quantidade média dessa substância está diferente da especificada. Para verificar essa suspeita, a indústria selecionou uma amostra aleatória de 40 comprimidos desse lote, observando uma quantidade média de ácido acetil salicílico igual a 5.2 gramas e um desvio padrão de 0.7 gramas.

## Exemplo 2

Uma indústria farmacêutica especifica que em certo analgésico a quantidade média de ácido acetil salicílico deve ser 5.5 gramas por comprimido. A indústria suspeita que houve problemas na produção de um determinado lote e que, nesse lote, a quantidade média dessa substância está diferente da especificada. Para verificar essa suspeita, a indústria selecionou uma amostra aleatória de 40 comprimidos desse lote, observando uma quantidade média de ácido acetil salicílico igual a 5.2 gramas e um desvio padrão de 0.7 gramas.

## Pergunta

Você tem informações suficientes para executar um teste formal de hipóteses?

Em caso afirmativo, formule a  $H_0$  e a  $H_1$ .

## Quais são as variáveis?

### Exemplo 2

Uma indústria farmacêutica especifica que em certo analgésico a quantidade média de ácido acetil salicílico deve ser 5.5 gramas por comprimido. A indústria suspeita que houve problemas na produção de um determinado lote e que, nesse lote, a quantidade média dessa substância está diferente da especificada. Para verificar essa suspeita, a indústria selecionou uma amostra aleatória de 40 comprimidos desse lote, observando uma quantidade média de ácido acetil salicílico igual a 5.2 gramas e um desvio padrão de 0.7 gramas.

### Modelo

variável dependente  $\sim$  variável independente

Significância

Felipe  
Figueiredo

Testes de  
Hipóteses

Hipóteses

Poder

Significância

O p-valor é...

Exercício

O p-valor não é...

Aprofundamento

## Quais são as variáveis?

### Exemplo 2

Uma indústria farmacêutica especifica que em certo analgésico a quantidade média de ácido acetil salicílico deve ser 5.5 gramas por comprimido. A indústria suspeita que houve problemas na produção de um determinado lote e que, nesse lote, a quantidade média dessa substância está diferente da especificada. Para verificar essa suspeita, a indústria selecionou uma amostra aleatória de 40 comprimidos desse lote, observando uma quantidade média de ácido acetil salicílico igual a 5.2 gramas e um desvio padrão de 0.7 gramas.

### Modelo

variável dependente  $\sim$  variável independente

peso de AAS  $\sim$  lote

Significância

Felipe  
Figueiredo

Testes de  
Hipóteses

Hipóteses

Poder

Significância

O p-valor é...

Exercício

O p-valor não é...

Aprofundamento

## Resposta

- Temos as informações necessárias para o teste

- Hipóteses

$$H_0 : \mu = 5.5$$

$$H_1 : \mu \neq 5.5$$

- Dados

$$n = 40, \bar{x} = 5.2, s = 0.7$$

## Resposta

- Temos as informações necessárias para o teste

- Hipóteses

$$H_0 : \mu = 5.5$$

$$H_1 : \mu \neq 5.5$$

- Dados

$$n = 40, \bar{x} = 5.2, s = 0.7$$

## Resultado (bruto)

$$p = 0.0099$$

### Interpretação

A probabilidade de observarmos **ao acaso** um valor **tão ou mais discrepante como 5.2 g** é 0.0099.

Como esta prob. é menor que o nível de significância estabelecido  $\alpha = 0.05$ , rejeitamos a hipótese de que a quantidade média é igual a 5.5 g ao nível de significância de 5%.

### Resultado

Significância

Felipe  
Figueiredo

Testes de  
Hipóteses

Hipóteses

Poder

Significância

O p-valor é...

Exercício

O p-valor não é...

Aprofundamento

## Interpretação

A probabilidade de observarmos **ao acaso** um valor **tão ou mais discrepante como 5.2 g** é 0.0099.

Como esta prob. é menor que o nível de significância estabelecido  $\alpha = 0.05$ , rejeitamos a hipótese de que a quantidade média é igual a 5.5 g ao nível de significância de 5%.

## Resultado

(...) a dose média de ácido acetil salicílico (...) por comprimido é 5.2 g e é significativamente diferente de 5.5 g ( $p = 0.0099$ ).

## Conclusão

Significância

Felipe  
Figueiredo

Testes de  
Hipóteses

Hipóteses

Poder

Significância

O p-valor é...

Exercício

O p-valor não é...

Aprofundamento



## Exemplo 2

### Interpretação

A probabilidade de observarmos **ao acaso** um valor **tão ou mais discrepante como 5.2 g** é 0.0099.

Como esta prob. é menor que o nível de significância estabelecido  $\alpha = 0.05$ , rejeitamos a hipótese de que a quantidade média é igual a 5.5 g ao nível de significância de 5%.

### Resultado

(...) a dose média de ácido acetil salicílico (...) por comprimido é 5.2 g e é significativamente diferente de 5.5 g ( $p = 0.0099$ ).

### Conclusão

O lote (...) está fora da especificação de 5.5 g (...) por comprimido, portanto o lote está reprovado.

Significância

Felipe  
Figueiredo

Testes de  
Hipóteses

Hipóteses

Poder

Significância

O p-valor é...

Exercício

O p-valor não é...

Aprofundamento

Outra hipótese, outra análise, outro resultado...



Significância

Felipe  
Figueiredo

Testes de  
Hipóteses

Hipóteses

Poder

Significância

O p-valor é...

Exercício

O p-valor não é...

Aprofundamento

## Observe que...

Se tivéssemos formulado as hipóteses que a média da amostra é maior que 5.5 g, qual você acha que seria o resultado?

Qual seria a conclusão neste caso?

- Nessa situação, podemos usar o intervalo de confiança para realizar o teste de hipóteses.
- Como queremos um teste a 5% de significância, calcularemos um intervalo de 95% de confiança.

### Lembre-se

$$\text{significância} + \text{confiança} = 1$$

#### Significância

Felipe  
Figueiredo

#### Testes de Hipóteses

Hipóteses

Poder

Significância

O p-valor é...

Exercício

O p-valor não é...

#### Aprofundamento

## Exemplo 2 (a revanche)

### IC da média

- Dados:  $n = 40$ ,  $\bar{x} = 5.2$ ,  $s = 0.7$
- IC:  $[4.976, 5.424] \approx [5.0, 5.4]$

#### Significância

Felipe  
Figueiredo

#### Testes de Hipóteses

Hipóteses

Poder

Significância

O p-valor é...

**Exercício**

O p-valor não é...

#### Aprofundamento

## Exemplo 2 (a revanche)

### IC da média

- Dados:  $n = 40$ ,  $\bar{x} = 5.2$ ,  $s = 0.7$
- IC:  $[4.976, 5.424] \approx [5.0, 5.4]$

### Resultado

A quantidade média neste lote (...) está entre 5.0 e 5.4 gramas, com 95% de confiança.

### Interpretação

A “meta” 5.5 g não está contida no IC.

Significância

Felipe  
Figueiredo

Testes de  
Hipóteses

Hipóteses

Poder

Significância

O p-valor é...

Exercício

O p-valor não é...

Aprofundamento

## Exemplo 2 (a revanche)

### IC da média

- Dados:  $n = 40$ ,  $\bar{x} = 5.2$ ,  $s = 0.7$
- IC:  $[4.976, 5.424] \approx [5.0, 5.4]$

### Resultado

A quantidade média neste lote (...) está entre 5.0 e 5.4 gramas, com 95% de confiança.

### Interpretação

A “meta” 5.5 g não está contida no IC.

### Conclusão

(...), portanto o lote está reprovado.

Significância

Felipe  
Figueiredo

Testes de  
Hipóteses

Hipóteses

Poder

Significância

O p-valor é...

Exercício

O p-valor não é...

Aprofundamento

## Exemplo 2 (a revanche)

### IC da diferença

- Dados:  $n = 40$ ,  $\bar{x} = 5.2$ ,  $s = 0.7$ ,  $\mu = 5.5$
- $\bar{d} = -0.300$ , IC:  $[-0.524, -0.076] \approx [-0.5, -0.1]$

### Resultado

A diferença entre a média desta amostra e a meta (...) está entre -0.5 e -0.1 gramas, com 95% de confiança.

### Interpretação

A “meta de igualdade”  $d = 0$  g não está contida no IC.

Significância

Felipe  
Figueiredo

Testes de  
Hipóteses

Hipóteses

Poder

Significância

O p-valor é...

Exercício

O p-valor não é...

Aprofundamento

## Exemplo 2 (a revanche)

### IC da diferença

- Dados:  $n = 40$ ,  $\bar{x} = 5.2$ ,  $s = 0.7$ ,  $\mu = 5.5$
- $\bar{d} = -0.300$ , IC:  $[-0.524, -0.076] \approx [-0.5, -0.1]$

### Resultado

A diferença entre a média desta amostra e a meta (...) está entre -0.5 e -0.1 gramas, com 95% de confiança.

### Interpretação

A “meta de igualdade”  $d = 0$  g não está contida no IC.

### Conclusão

(...), portanto o lote está reprovado.

Significância

Felipe  
Figueiredo

Testes de  
Hipóteses

Hipóteses

Poder

Significância

O p-valor é...

Exercício

O p-valor não é...

Aprofundamento



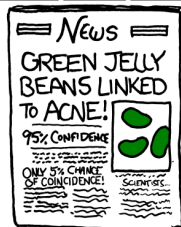
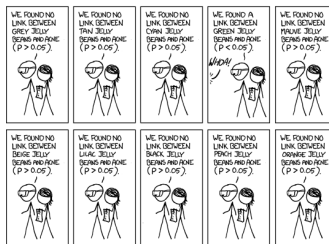
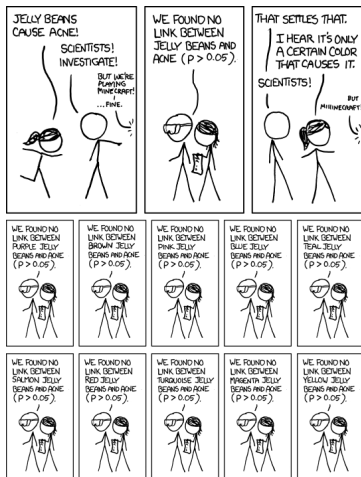
## 1 Testes de Hipóteses

- Hipóteses
- Poder estatístico
- Significância Estatística
- O p-valor é...
- Exercício
- O p-valor não é...

## 2 Aprofundamento

- Aprofundamento

# Como a Ciência Médica é vista na mídia



Fonte: <https://xkcd.com/882/>

## Significância

Felipe  
Figueiredo

## Testes de Hipóteses

Hipóteses

Poder

Significância

O p-valor é...

Exercício

O p-valor não é...

## Aprofundamento

# Why Most Published Research Findings Are False

John P. A. Ioannidis

Published: August 30, 2005 • <http://dx.doi.org/10.1371/journal.pmed.0020124>

Article	Authors	Metrics	Comments	Related Content
				

## Abstract

Modeling the Framework  
for False Positive  
Findings

Bias

Testing by Several  
Independent Teams

## Abstract

### Summary

There is increasing concern that most current published research findings are false. The probability that a research claim is true may depend on study power and bias, the number of

DOREY, F. 2010 Clin Orthop Relat Res.

*"The concept of a p value is not simple and any statements associated with it must be considered cautiously."*

## Estes são erros comuns de interpretação

### O p-valor assume que...

- 1 a hipótese nula é **verdadeira**
- 2 a **única** causa da diferença observada é devida ao acaso

Significância

Felipe  
Figueiredo

Testes de  
Hipóteses

Hipóteses

Poder

Significância

O p-valor é...

Exercício

O p-valor não é...

Aprofundamento

Estes são erros comuns de interpretação

### O p-valor assume que...

- 1 a hipótese nula é verdadeira
- 2 a única causa da diferença observada é devida ao acaso

### Portanto o p-valor **não é**

- a probabilidade de que a hipótese nula seja verdadeira
- a probabilidade de que a diferença observada seja devido ao acaso

Significância

Felipe  
Figueiredo

Testes de  
Hipóteses

Hipóteses

Poder

Significância

O p-valor é...

Exercício

O p-valor não é...

Aprofundamento

Estes são erros comuns de interpretação

Significância

Felipe  
Figueiredo

Testes de  
Hipóteses

Hipóteses

Poder

Significância

O p-valor é...

Exercício

O p-valor não é...

Aprofundamento

### O p-valor assume que...

- 1 a hipótese nula é verdadeira
- 2 a única causa da diferença observada é devida ao acaso

### Portanto o p-valor **não é**

- a probabilidade de que a hipótese nula seja verdadeira
- a probabilidade de que a diferença observada seja devido ao acaso

O p-valor não pode ser usado para concluir suas próprias premissas.

## A questão importante aqui é:

MESMO SE a hipótese nula **for verdadeira**, ainda assim você pode observar (ao acaso) uma diferença como resultado do experimento.

(ex., muita variabilidade, amostras pequenas, etc.).

**Isso** é o erro tipo I. Trabalhamos para que isso seja raro (não mais que 5% das vezes).



- 1 Testes de Hipóteses
  - Hipóteses
  - Poder estatístico
  - Significância Estatística
  - O p-valor é...
  - Exercício
  - O p-valor não é...

- 2 Aprofundamento
  - Aprofundamento

## Leitura obrigatória

- Capítulo 10.
- Capítulo 11.

## Leitura recomendada

### Para entender melhor Poder Estatístico

- Capítulo 15 (até seção **Probabilidade de obter um falso positivo no lab [...]**)
- Cap 22 (seção: **Interpretando uma afirmação sobre tamanho amostral e poder**)
- Motulsky, (2018) chap 19, **Interpreting a Result That Is Not Statistically Significant** (disponível gratuitamente online)
- Dorey, F (2010) **In Brief: The P Value: What Is It and What Does It Tell You?**
- Gardner, MJ; Altman, DG (1986) **Confidence intervals rather than P values: estimation rather than hypothesis testing.**