# GEDS - Grupo de Estudos de Data Science

# Semana 8 Teste de Hipótese

"Um dos problemas a serem resolvidos pela Inferência Estatística é o de testar uma hipótese. Isto é, feita determinada afirmação sobre uma população, usualmente sobre um parâmetro<sup>1</sup> dessa, desejamos saber se os resultados experimentais provenientes de uma amostra contrariam ou não tal afirmação.

O objetivo do teste estatístico de hipóteses é, então, fornecer uma metodologia que nos permita verificar se os dados amostrais trazem evidências que apóiem ou não uma hipótese (estatística) formulada."

Livro Estatística Básica - Morettin e Bussab - Capítulo 12.

#### Links úteis:

#### Teste de Hipótese - Khan Academy

https://pt.khanacademy.org/math/statistics-probability/significance-tests-one-sample/idea-of-significance-tests/v/simple-hypothesis-testing

Sequência de vídeos sobre Teste de Hipótese - Canal Professor Guru 4º video é o passo a passo de como formular um teste de Hipótese <a href="https://www.youtube.com/playlist?list=PL7xT0Gz6G0-TfV-S6WiGDvIsZds6Pv\_q8">https://www.youtube.com/playlist?list=PL7xT0Gz6G0-TfV-S6WiGDvIsZds6Pv\_q8</a>

#### Material UFSC - bem completo!

https://www.inf.ufsc.br/~andre.zibetti/probabilidade/teste-de-hipoteses.html#testes\_de\_hip% C3%B3teses\_estat%C3%ADsticas

#### Introdução à Testes de Hipótese



Pergunta: Qual o tempo médio da sua casa até o trabalho?

tempo médio = 1h30min

Passo 1: Formular a Hipótese

H0: hipótese nula - o que eu quero provar! H1: hipótese alternativa - complementar de H0

H0: tempo médio é igual 1h30min

H1: tempo médio é diferente de 1h30min

<sup>&</sup>lt;sup>1</sup>Uma característica matemática de uma determinada população, por exemplo, as medidas de centralidade.

### pyladies. São Paulo

#### GEDS - Grupo de Estudos de Data Science

#### Passo 2: Dados para testar

Amostra de dias em que a Maria foi para o trabalho.

Possíveis viéses da amostra: medir em dias de chuva, finais de semana, dias de rodízio do carro, dias que ela foi trabalho em horários alternativos, sextas-feiras, vésperas de feriado.

#### Passo 3: Tipos de Erro

Ao tomar uma decisão podemos cometer algum erro. Chamaremos esses erros de erro do tipo I e II.

Erro do Tipo I: Rejeitar a hipótese nula quando ela é verdadeira;

Erro do Tipo II: Não Rejeitar a hipótese nula quando ela é falsa.

Decisão	Se $H_0$ é verdadeira	Se $H_0$ é falsa
Rejeitar $H_{f 0}$	Erro do Tipo I	Nenhum erro
Não Rejeitar $H_{f 0}$	Nenhum erro	Erro do Tipo II

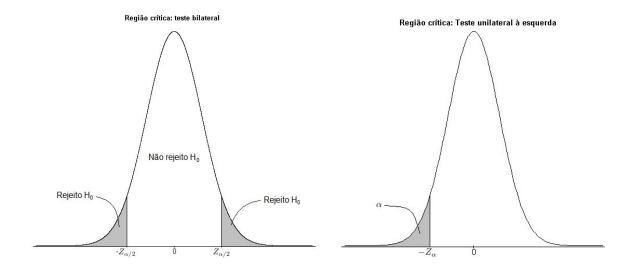
#### Passo 4: Probabilidade de cometer um erro

Nível de significância: 5%

Por exemplo, a cada 100 vezes eu acerto 95, ou seja, a cada 100 idas ao trabalho 95 delas será dentro da hipótese testada.

#### Passo 5: Construção da estatística de teste

Na hipótese, tempo médio é 90min. Com uma amostra é possível calcular a média amostral.



## pyladies. São Paulo

#### GEDS - Grupo de Estudos de Data Science

Nós usamos valores-p para tirar conclusões em testes de significância. Mais especificamente, nós comparamos o valor-p com um nível de significância  $\alpha$  para tirar conclusões sobre nossas hipóteses.

Precisamos saber o limite superior e o limite inferior para saber se nós rejeitarmos a nossa hipótese (hipótese nula) e utilizaremos a hipótese alternativa.

Z(alfa)/2 => é o que coloca para fora (rejeição)

Se o valor-p for menor que o nível de significância escolhido, nós rejeitamos a hipótese nula H0 em favor da hipótese alternativa Ha . Se o valor-p for maior ou igual ao nível de significância, então não rejeitamos a hipótese nula H0. Isso não significa que nós aceitamos H0. Para resumir:

valor-
$$p < \alpha \Rightarrow \text{rejeitar } H_0 \Rightarrow \text{aceitar } H_a$$

valor-
$$p \geq \alpha \Rightarrow$$
 não rejeitar  $H_0$