



Data Science Academy

www.datascienceacademy.com.br

Microsoft Power BI Para Data Science

Teste de Hipótese



Vamos compreender o que é Teste de Hipótese, um dos conceitos fundamentais em Estatística Inferencial.

No mundo da **Estatística**, uma **hipótese** é uma **suposição** sobre um **parâmetro** específico de uma população, tal como média, proporção ou desvio padrão. Podemos fazer uma suposição sobre o valor de um parâmetro da população, coletar uma amostra desta população, medir a amostra e atestar se a amostra suporta ou não a nossa suposição. O mundo dos negócios é cheio de exemplos que lidam com testes de hipóteses:

- 1- Uma indústria de lâmpadas que fabrica lâmpadas fluorescentes afirma que suas lâmpadas usam 75% menos energia e duram 10 vezes mais que as lâmpadas comuns. Um laboratório independente pode testar esta afirmação com um **teste de hipótese**.
- 2- Um artigo recente de um grande jornal, afirmou que o excesso de tempo em redes sociais poderia afetar a capacidade intelectual das pessoas. Um pesquisador poderia validar esta afirmação usando um teste de hipótese.
- 3- Antes da crise, os bancos cobravam em média R\$40 em taxas administrativas de contas correntes de pessoas físicas. Um órgão regulador do governo poderia testar a hipótese de os bancos estarem cobrando mais de R\$40 em taxas.

Teste de Hipótese é um dos procedimentos estatísticos mais usados atualmente. Cuidados especiais devem ser tomados para definir propriamente o teste e para gerar as conclusões.

O objetivo do teste de hipótese é decidir se determinada afirmação sobre um parâmetro populacional é ou não apoiada pela evidência obtida de dados amostrais. Cada teste de hipótese tem uma hipótese nula e uma hipótese alternativa, representados por:

H_0 - Hipótese nula

H_A - Hipótese alternativa

A hipótese nula representa o status quo, ou seja, comprovar uma suposição ou afirmação. A **hipótese nula** é para ser verdadeira, a menos que seja comprovada por uma evidência contrária. A **hipótese alternativa** representa o oposto da hipótese nula. Você precisa ser **cuidadoso** ao definir a **hipótese nula** e a **hipótese alternativa**. Vamos a um exemplo de como definir as hipóteses.



Neste exemplo, nós estamos assumindo que os usuários de internet ficam em média 56 segundos em uma web page.

Suponhamos que o propósito do teste seja determinar se a média da população é igual a um valor específico. Definiríamos assim nossas hipóteses:

$H_0: \mu = 56$ segundos (status quo)

$H_0 = 56$ segundos

A hipótese alternativa reflete a condição oposta, ou seja, que o tempo médio que os internautas ficam em uma web page **não** é igual a 56 segundos.

$H_A: \mu \neq 56$ segundos

$H_A \neq 56$ segundos

Uma declaração de hipótese pode somente ser usada com parâmetro da população (**tal como μ**), não com uma estatística de amostra (**tal como a média da amostra**). O propósito do **teste de hipótese** é **atestar** uma **conclusão** sobre os parâmetros da população, sobre os quais **não** temos conhecimento completo. Mais um exemplo:

Vamos imaginar um fabricante de lâmpadas que afirma que desenvolveu um novo produto cujo tempo médio de vida **supera a média da indústria de 8.000 horas**.



Para testarmos esta afirmação, definimos o seguinte teste de hipótese:

- $H_0: \mu \leq 8.000$ horas
- $H_A: \mu > 8.000$ horas

Perceba que a hipótese alternativa foi usada para representar a afirmação feita pelo fabricante. As **8.000** horas de tempo de vida em **média**, é considerado ser verdadeiro (status quo) e por isso foi atribuído ao **teste de hipótese nula**. Definir as hipóteses nulas e alternativas, depende da natureza do teste e da pessoa que o está conduzindo. Uma vez definidas as hipóteses, realizamos experimentos, coletamos as estatísticas e rejeitamos ou não a hipótese. Mais alguns exemplos:

Definir as hipóteses nulas e alternativas, depende da natureza do teste e da pessoa que o está conduzindo

$$H_0: \mu = 1.8$$
$$H_A: \mu > 1.8$$

Este teste seria usado por alguém que acredita que o uso de banda larga aumentou e quer suportar que a média de uso de banda larga é agora maior que 1.8 GB por mês.

$$H_0: \mu = 1.8$$
$$H_A: \mu < 1.8$$

Este teste seria usado por alguém que acredita que o uso de banda larga diminuiu e quer suportar que a média de uso de banda larga é agora menor que 1.8 GB por mês.

$$H_0: \mu = 1.8$$
$$H_A: \mu \neq 1.8$$

Este teste seria usado por alguém que não possui uma expectativa específica, mas quer testar a suposição que a média de uso de banda larga é 1.8 GB por mês.

A hipótese nula H_0 representa o status quo, ou seja, a circunstância que está sendo testada e o objetivo dos testes de hipóteses é sempre tentar **rejeitar** a hipótese nula. A hipótese alternativa H_A representa o que se deseja provar ou estabelecer, sendo formulada para **contradizer** a hipótese nula. Existem apenas **2** afirmações que podemos fazer sobre a **hipótese nula**:

Rejeitar

Não Rejeitar



A razão pela qual estamos limitados a apenas 2 conclusões possíveis, é que o teste de hipótese se baseia em “provar contradições”. Com isso, nós podemos apenas concluir que a *hipótese pode ser verdadeira*, **mas não temos evidências suficientes para afirmar que a hipótese nula é realmente verdadeira**. Por conta desta limitação, NUNCA podemos aceitar a hipótese nula. Podemos apenas dizer que: **não há evidências suficientes para rejeitar a hipótese nula**.

No curso Business Analytics aqui na DSA, os alunos aprendem a criar Testes de Hipótese usando linguagem R. Os Testes de Hipótese são amplamente utilizados no mundo dos negócios, em Digital Marketing através de Testes A/B, em análise e detecção de fraudes, em análise de redes sociais e muito mais.