

aulas.descomplica.com.br /pos/pos-graduacao-em-data-science/turma/statistics-for-data-science-4cc58c/aula/teste-de-hip...

# Teste de hipótese

---

## Introdução

Experimentação tem sido um assunto muito importante para empresas que entendem o valor de tomar decisões baseadas em dados. Nesse sentido, testes de hipótese possuem um papel central nas tomadas de decisão. Nesse módulo, veremos quais testes podemos usar e como utilizar os seus resultados, além de entender quais as vantagens e o que é necessário para uma experimentação em empresas de tecnologia.

## Objetivos da aula

- O que é um teste de hipótese
- Teste de Médias
- Teste de Proporções
- Experimentação Online

## Resumo

Um teste de hipótese tem como objetivo decidir se uma hipótese feita sobre um parâmetro populacional é ou não verdadeira, e usamos dados amostrais de um experimento dessa hipótese para tomar essa decisão. Portanto, o teste de hipótese é uma ferramenta que possibilita uma tomada de decisão baseada em dados.

No teste, vamos definir duas hipóteses: a hipótese nula e a hipótese alternativa.

- Hipótese Nula ( $H_0$ ): é uma afirmação a respeito do valor do parâmetro populacional que deve ser testada.
- Hipótese Alternativa ( $H_1$ ): é uma afirmação a respeito do parâmetro que aceitaremos como provavelmente verdadeiro caso  $H_0$  seja rejeitada.

Por exemplo, digamos que você queira entender qual canal de vendas geral tem uma receita média maior, a do site ou a do app. As hipóteses ficariam:

$H_0$ : média de vendas no site = média de vendas no app

$H_1$ : média de vendas no site  $\neq$  média de vendas no app

Ao fim do teste, depois de executar todos os testes estatísticos e escolhermos uma das hipóteses como verdadeira, existe uma chance de estarmos cometendo um erro. Isso porque os dados que temos são amostrais e sempre vai existir uma chance de que os resultados tenham sido resultado do acaso, e não devido a uma diferença real nos dados. Assim, é importante que saibamos os riscos de tomar uma decisão

no teste de hipótese. Existem dois tipos de erros: Erro do Tipo I e Erro do Tipo II.

O Erro do Tipo I é a probabilidade de se rejeitar  $H_0$  quando na verdade  $H_0$  é verdadeira. Esse é o erro mais utilizado nos testes de hipótese. Na prática, determinamos um valor máximo de tolerância (nível de significância), que costuma variar entre 1% e 5% e, se a probabilidade do erro do tipo I for maior do que o nível de significância, então não rejeitamos a hipótese nula, ou seja, podemos aceitá-la como verdade.

O Erro do Tipo II é a probabilidade de se rejeitar  $H_1$  quando na verdade  $H_1$  é verdadeira. É importante principalmente quando os resultados estão muito próximos, por exemplo, a média de vendas no site foi muito próxima da média de vendas no app. Na prática, geralmente queremos que o erro do tipo II seja menor do que 20%.

O Teorema do Limite Central nos diz que, para uma amostra grande o suficiente, a média tem distribuição Normal. Um resultado disso é o teste de hipótese entre médias! Vamos ver um passo a passo para entendermos como ele funciona.

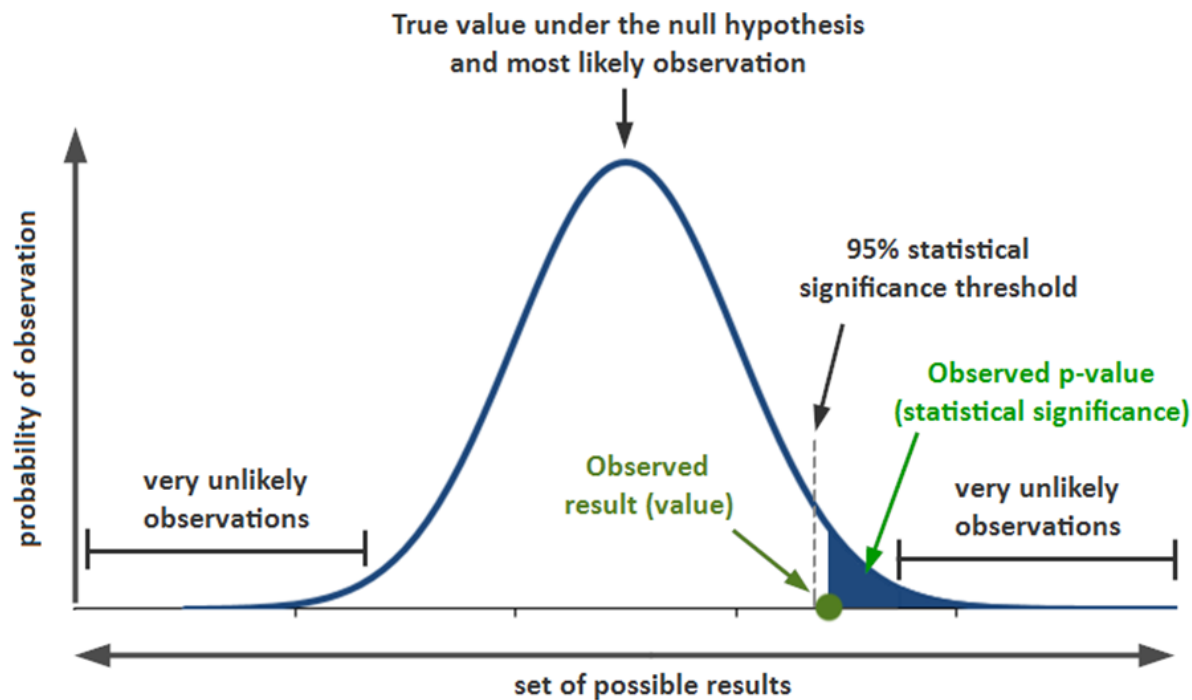
1. Pelo teorema do limite central, a média tem distribuição normal
2. Num teste de hipótese de comparação entre duas médias, temos duas hipóteses: a nula e a alternativa. A hipótese nula vai dizer que as duas médias são iguais, ela basicamente diz que a ideia que queremos testar não altera a realidade atual e, por isso, as duas médias são iguais. A hipótese alternativa diz que elas são diferentes, ou seja, a ideia que testamos alterou a realidade atual, o status quo foi mudado.
3. A amostra obtida sob a realidade atual vai ser chamada de amostra controle, ou grupo controle. A amostra obtida sob a realidade nova, com a ideia que está sendo testada, será chamada de amostra variante, ou grupo variante.



Fonte: <https://usabilitygeek.com/wp-content/uploads/2020/05/mmm.jpg> (acesso em 21/12/2022)

4. Para medir se houve diferença entre as médias dos grupos controle e variante, vamos diminuir a média do grupo controle pela média do grupo variante, formando uma nova variável. Vamos chamá-la de  $X$ .
5. Essa nova variável  $X$  tem distribuição Normal!
6. Se considerarmos que  $H_0$  é verdadeiro, então  $X$  tem média igual a 0.
7. Se soubermos quanto vale a variância da população, usamos a distribuição Normal para o teste. Se não, usamos a distribuição T de Student.
8. Olhamos para o valor observado das médias das amostras controle e variante, bem como os tamanhos das amostras e suas variâncias, e chegamos a um valor que chamamos de estatística de teste.
9. Se a estatística de teste for muito improvável de ocorrer segundo a distribuição de  $X$  (diferença das médias) sob  $H_0$  (média de  $X$  é 0), então rejeitamos a hipótese nula e dizemos que as médias dos grupos são diferentes. Se não, dizemos que são iguais.
10. Chamamos de p-valor a probabilidade de ocorrência da estatística de teste ou valores mais extremos que ela, segundo a distribuição de  $X$ , sob  $H_0$ . Ou seja, o p-valor é a probabilidade de valores iguais ou mais extremos do que a estatística de teste ocorrerem. O p-valor é a probabilidade estimada do

erro do tipo I, calculada usando os dados amostrais.



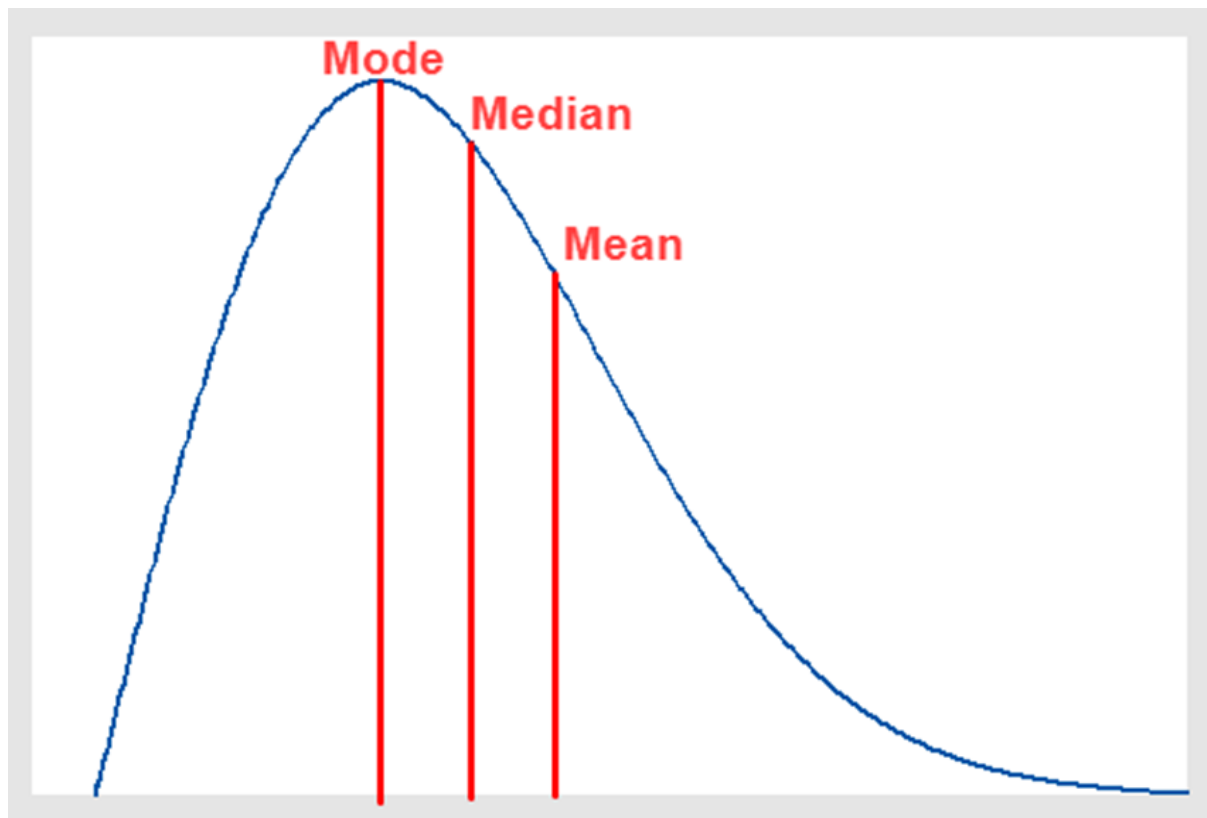
Fonte: <https://www.simplypsychology.org/p-value.png> (acesso em 21/12/2022)

11. Na prática, se o p-valor obtido for baixo, nós rejeitamos  $H_0$  e dizemos que as médias dos grupos são diferentes. Se o p-valor for alto, nós não rejeitamos  $H_0$  e dizemos que a média dos grupos são iguais.
12. Para decidir se um p-valor é baixo, definimos antes do teste começar um threshold, a quem chamaremos de nível de significância. Se o p-valor for menor ou igual ao nível de significância, então ele é baixo.

Por exemplo, digamos que o nosso nível de significância é 5%. Sob a hipótese nula, as médias propostas são iguais. Daí,

- Se o p-valor for menor do que 5%, rejeitamos a hipótese nula com 95% de confiança;
- Se o p-valor for menor do que 1%, rejeitamos a hipótese nula com 99% de confiança;
- Senão, não rejeitamos a hipótese nula.

Vale dizer que o teste de médias pode ser utilizado para médias de qualquer tipo de dados, mesmo dados que não possuem distribuição Normal. Isso graças ao Teorema do Limite Central! Entretanto, não é prudente utilizá-lo para médias de dados com distribuição muito assimétricas. Isso porque a média não vai ser uma boa medida de tendência central, pois não vai representar bem nenhuma faixa dos seus dados. Então, apesar de não ser errado conduzir um teste de médias, não seria a melhor solução.



Fonte: [https://statisticsbyjim.com/wp-content/uploads/2021/10/Right\\_skewed\\_central\\_tendency.png](https://statisticsbyjim.com/wp-content/uploads/2021/10/Right_skewed_central_tendency.png) (acesso em 21/12/2022)

Pensando no teste de hipótese para proporção, se considerarmos que cada observação segue uma distribuição de Bernoulli, onde 0 representa um fracasso e 1 representa sucesso, a média dessas observações é uma proporção! Portanto, a proporção de sucessos é em si uma média e, por isso, todos os passos acima são válidos também para a proporção.

A grande particularidade do teste de proporções é que, como consideramos que cada observação é uma Bernoulli, então sabemos exatamente qual é a sua variância -  $p*(1-p)$  -, o que simplifica o nosso teste.

E quais os tipos de experimentos fazemos no dia a dia de mercado? Em geral, experimentos online! Um experimento nada mais é do que um teste científico, que envolve descobrir o que esperaríamos observar se uma ideia estivesse correta, e comparar essa expectativa com o que realmente observamos.

Quando falamos de experimentação, é muito importante entender a diferença entre correlação e causalidade, e qual o nível de confiabilidade de cada tipo de experimento. A correlação é uma medida de afinidade entre duas métricas, Causalidade é uma ligação de causa e efeito entre dois eventos. Mas veja, a correlação não implica causalidade, ou seja, uma correlação alta entre duas variáveis não implica que uma variável tem uma causa no efeito de mudança na outra.

Nesse sentido, a experimentação é a forma mais confiável de provarmos a causalidade. Se queremos provar que uma mudança no fluxo de checkout de vendas de um e-commerce muda o número de vendas finalizadas, a melhor forma de provar isso seria conduzindo um experimento.

Na prática, essas são as etapas de condução de um experimento:

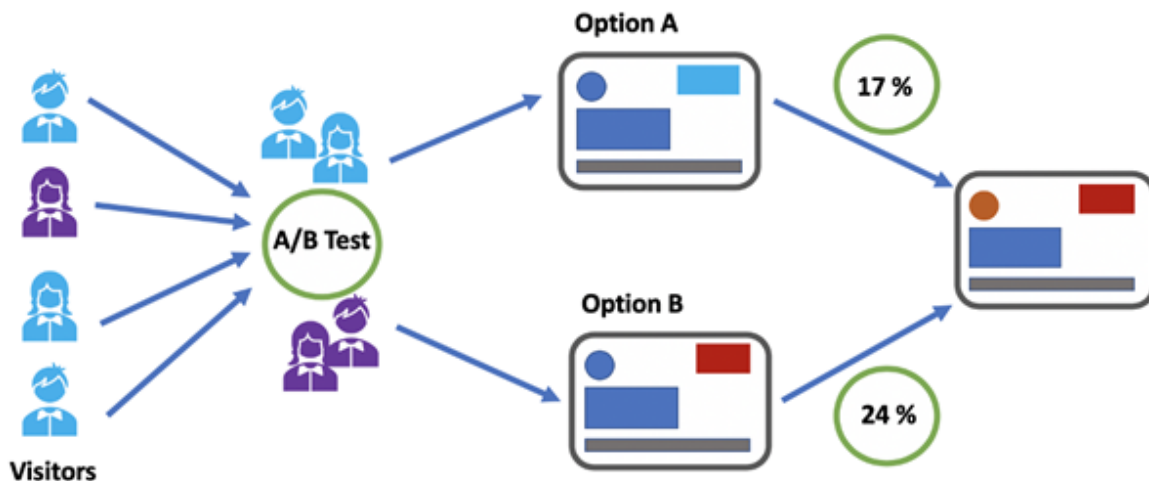


Fonte: <https://medium.com/1point96/the-experimentation-process-has-been-overly-simplified-a5845aabc8ac> (acesso em 21/12/2022)

1. Define um objetivo
2. Define a ideia que será testada
3. Cria os grupos controle e variante
4. Executa o experimento
5. Analisa os resultados
6. Se a ideia der bons resultados, aplicamos para todos os usuários. Se não, voltamos ao passo 1.

Em geral, para executar um experimento, precisamos de uma ferramenta (interna ou de mercado) que aleatoriza os usuários entre o grupo controle e variante. O grupo controle tem uma experiência igual à que já vinha tendo, ou seja, nada muda para o grupo controle. Já o grupo variante tem uma experiência

exatamente igual ao grupo controle, exceto pela hipótese (ideia) que está sendo testada.



Fonte: <https://towardsdatascience.com/how-to-conduct-a-b-testing-3076074a8458> (acesso em 21/12/2022)

### Como aplicar na prática o que aprendeu

Para praticar o que você aprendeu, procure o time de Experimentação da sua empresa. Se ele não existe, procure saber como fazer testes A/B por aí, ou se alguém já fez algo do tipo. Se não é algo comum, essa pode ser uma ótima oportunidade para você liderar essa pauta. Essa é uma grande diferença competitiva para as empresas, e quanto antes você começar a experimentar, melhor!

Além disso, você pode utilizar dados já disponíveis para você e aplicar os testes que aprendemos nas videoaulas, usando os notebooks em python disponibilizados.

### Conteúdo bônus

#### Tópicos avançados

Para complementar seus estudos, vale a pena a leitura do livro *Experimentation Works: The Surprising Power of Business Experiments*, e do livro *Trustworthy Online Controlled Experiments: A Practical Guide to A/B Testing*.

Além disso, recomendo assistir a esse [vídeo](#) de introdução sobre experimentação na Netflix. Pode ser interessante pesquisar sobre Inferência Causal também, um tema que acaba se confundindo com o de Experimentação.

Caro estudante, você consegue acessar os códigos utilizados na disciplina no link a seguir: <https://github.com/FaculdadeDescomplica/Statistics-for-Data-Science>

### Referência Bibliográfica

**How to conduct A/B Testing?**. Disponível em: <<https://towardsdatascience.com/how-to-conduct-a-b-testing-3076074a8458>>. (Acesso em 21/12/2022)

**Online Controlled Experiments: Lessons from Running A/B/n Tests for 12 Years**. Disponível em: <<https://dl.acm.org/doi/abs/10.1145/2783258.2785464>>. (Acesso em 21/12/2022)