

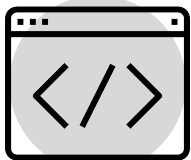
# Profissão: Cientista de Dados



# BOAS PRÁTICAS



# Inferência



- **Lógica indutiva vs dedutiva**
- **O problema da inferência**
- **Teste de significância**
- **Teste de hipóteses no Python**
- **O poder do teste**



# Lógica indutiva vs dedutiva

- Entenda a diferença entre pensamento dedutivo e indutivo: O pensamento dedutivo parte de uma premissa geral para chegar a uma conclusão específica, enquanto o pensamento indutivo parte de um fato particular para chegar a uma conclusão geral. Ambos são importantes na ciência de dados e devem ser usados de acordo com a situação.
- A inferência estatística é uma ferramenta crucial na ciência de dados, pois permite extrapolar conclusões de uma amostra para uma população com um certo nível de confiança. Use-a para fazer previsões e tirar conclusões.



# O problema da inferência

- Sempre considere o problema básico da inferência estatística: você está tentando estimar características de uma população maior com base em uma amostra. Mantenha isso em mente ao analisar seus dados e tirar conclusões.
- Aplique problemas clássicos de inferência estatística a situações reais. Por exemplo, o problema das duas urnas pode ser aplicado para determinar de qual fornecedor veio uma caixa de parafusos, com base em uma amostra de parafusos da caixa.
- Lembre-se de que os dados podem vir de um processo estocástico ou de uma amostra fixa e finita. Em ambos os casos, a inferência estatística é usada para analisar uma amostra da população e extrapolar conclusões sobre a população em geral.
- Prepare-se para a próxima etapa da inferência estatística: o teste de hipóteses. Este é um método importante para determinar se as diferenças observadas em seus dados são significativas ou se podem ser atribuídas ao acaso.



# Teste de significância

- Entenda a lógica indutiva: A lógica indutiva é usada para determinar se uma afirmação é provável com base em evidências. No exemplo da aula, o fabricante de calçados usou a lógica indutiva para determinar se o novo material é mais resistente.
- Esteja ciente do erro do tipo um: O erro do tipo um é a probabilidade de rejeitar erroneamente a hipótese nula quando ela é verdadeira. A estatística permite quantificar esse erro e definir um nível de erro que é considerado aceitável.
- Defina claramente a hipótese nula e alternativa: A hipótese nula é a afirmação que você está tentando refutar, enquanto a hipótese alternativa é a afirmação que você está tentando provar.
- Entenda o conceito de região crítica: A região crítica é a área sob a curva de distribuição onde a hipótese nula seria rejeitada. A probabilidade de observar uma estatística de teste dentro dessa região é usada para tomar uma decisão sobre a hipótese nula.



# Teste de hipóteses no Python

- Lembre-se de que o erro padrão diminui à medida que o tamanho da amostra aumenta. Portanto, sempre que possível, tente coletar uma amostra maior para tornar seu teste mais poderoso.
- Ao apresentar os resultados de um teste de hipóteses, explique claramente o conceito de região crítica e valor-p para que os resultados sejam compreendidos corretamente.
- Ao interpretar os resultados de um teste de hipóteses, não se esqueça de considerar o valor-p. Um valor-p menor indica uma evidência mais forte de que a hipótese nula é falsa.



# Poder do teste

- Ao realizar um estudo de tamanho amostral, é importante entender a diferença entre erro tipo um e erro tipo dois. O erro tipo um é a probabilidade de rejeitar a hipótese nula ( $H_0$ ) quando ela é verdadeira, enquanto o erro tipo dois é a probabilidade de não rejeitar  $H_0$  quando uma hipótese alternativa é verdadeira.
- A hipótese alternativa é geralmente baseada no conceito de uma diferença significativa. O que é considerado significativo é subjetivo e depende muito do conhecimento do negócio.
- O poder do teste é definido como um menos o erro do tipo dois. Quando se diz que um teste é mais poderoso, significa que a probabilidade de não rejeitar  $H_0$ , dada uma hipótese alternativa, é pequena.
- É importante entender a diferença entre significativo e significante. Significante é o que o teste de hipóteses consegue identificar, enquanto significativo é se a diferença tem um significado de negócio.





# Bons estudos!

