

# Ciência Reprodutível para Experimentos em Computação de Alto Desempenho

---

Pedro Bruel  
*phrb@ime.usp.br*  
April 29, 2021

# Introdução

---

# Roteiro

O que é Ciência Reprodutível?

Desafios para se fazer Ciência Reprodutível

Mão na Massa: Ferramentas para Reprodutibilidade

## Dependências e outros Recursos

- Jupyter Notebook, R, pacotes R, (docker?)
- Site com instruções e mais recursos (livros, ...)

## O que é Ciência Reprodutível?

---

# O que é Ciência Reprodutível?

# Crise de Reprodutibilidade

## **Desafios para se fazer Ciência Reprodutível**

---



# Computação de Alto Desempenho: Desafios

- Arquiteturas Específicas
  - Custo elevado
  - Como acessar?
- Software complexo
  - Dependências
- Consumo de Energia

# Computação de Alto Desempenho: Abordagens

- Virtualização
- Controle de versão
- Hospedagem pública
- Gerenciadores de dependências
- Simuladores

## Ciência Fechada: Desafios

- Como acessar publicações?
- Como compartilhar resultados?
- Software livre
- Análises de dados, figuras: código

## Ciência Fechada: Abordagens

- Documentos computacionais
- Arquivamento
- Gerenciadores de dependências
- Controle de versão
- Hospedagem pública

# Estatística: Desafios

- Datasaurus Dozen
  - <https://cran.r-project.org/web/packages/datasauRus/>
- O que é incerteza?
- Como planejar experimentos?
- Como analisar resultados?
  - O que é Machine Learning?

# Estatística: Abordagens

- Documentos computacionais
- Desenho de Experimentos
- Controle de versão
- Gráficos, antes de qualquer análise
- Fazer análises mais simples primeiro: mas fáceis de interpretar

# É Possível Garantir a Reprodutibilidade?

## **Mão na Massa: Ferramentas para Reprodutibilidade**

---



# Controle de Versão

# Hospedagem





# Arquivamento

## Conclusão

---

**É possível fazer Ciência (mais) Reprodutível!**

# Ciência Reprodutível para Experimentos em Computação de Alto Desempenho

---

Pedro Bruel  
*phrb@ime.usp.br*  
April 29, 2021