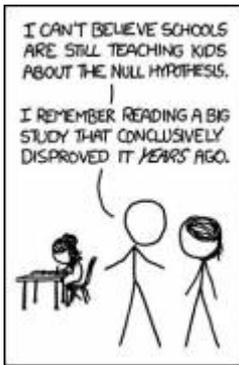


# Teste de Hipóteses

Antes de iniciar o roteiro anote o seu nome, altura (em centímetros e sem decimais) e sexo (M, F) na planilha [altura alunos](#).



O objetivo desse roteiro é auxiliar na compreensão de dois conceitos fundamentais do teste de hipótese estatística. O primeiro é entender o significado do  $p$ , ou a probabilidade de que, num cenário na qual a hipótese nula seja verdadeira, uma estatística de interesse tomada a partir de uma amostra seja igual ou mais extrema do que o valor observado. Dessa forma, o valor de  $p$  é um indicador da incompatibilidade entre o conjunto de dados coletados e a hipótese nula. Suponha um valor de  $p$  muito pequeno (próximo de 0,001). Isso significa que é pouco provável que a diferença observada (estatística de interesse) tenha sido obtida caso a hipótese nula fosse verdadeira (alta incompatibilidade). Diante de um resultado pouco provável, rejeitamos a hipótese nula e ficamos com a hipótese alternativa que pressupõe algo mais, além da variação aleatória. Se por outro lado, o valor de  $p$  é próximo de 1,00, significa que a diferença observada tem alta probabilidade de ser gerada a partir do cenário nulo, ou seja, de que a diferença observada seja atribuída apenas à variação aleatória decorrente do acesso limitado aos dados (alta compatibilidade). Mas qual deve ser o ponto de corte, ou seja, a partir de que valor crítico de  $P$  devemos aceitar ou rejeitar a hipótese nula? Convencionou-se arbitrariamente, em muitas áreas da ciência, que tal valor crítico é de 0.05 ou 5%.

O segundo objetivo desse roteiro é definir quando devemos aplicar um teste unicaudal ou bicaudal e como isso se relaciona à nossa hipótese estatística e à nossa hipótese científica. É muito importante que, além de fazer o roteiro, você discuta com colegas e professores os conceitos apresentados.

## Instalando o Rsampling

Para rodar esse roteiro utilizaremos duas ferramentas para o ensino de estatística desenvolvida por pessoas ligadas ao departamento de ecologia do IBUSP. O primeiro é o pacote **Rsampling** que contém instrumentação de permutação e aleatorização para a linguagem R. O outro é uma interface gráfica (**Rsampling-shiny**) para usar as ferramentas deste pacote no seu navegador. Não há necessidade de conhecimento prévio em R para fazer esse roteiro.

### Rsampling-shiny

O primeiro passo é baixar a interface gráfica e descompactar o arquivo em um diretório conhecido no seu computador. Sugerimos que crie uma pasta para a disciplina "planeco" dentro da pasta de documentos do computador que está trabalhando.

Baixe, para essa pasta, o arquivo "source code" *zip* ou *tar.gz*<sup>1</sup> da versão mais recente (1.7.0) do

Rsampling-shiny no link:

- [Versões Rsampling-shiny](#)

Descompacte o arquivo na mesma pasta.

*Depois de descompactar e testar o funcionamento (com os passos abaixo), você pode deletar o arquivo .zip ou .tar.gz*

## Usando o R

Abra o R pelo ícone  do programa <sup>2)</sup> instalado em seu computador. Usando o menu **Arquivo > mudar dir...** mude o diretório de trabalho para o diretório (i.e. a pasta) em que você salvou o Rsampling.

## Instalando os pacotes

Em seguida copie e cole a seguinte linha de comando no terminal do R para instalar os pacotes associados:

```
install.packages(c("Rsampling", "shiny", "PerformanceAnalytics"))
```

A seguinte mensagem irá aparecer no terminal do R “— Please select a CRAN mirror for use in this session —” e uma janela irá se abrir para que selecione o repositório mais próximo. Selecione **Brazil(SP 1)**. Aguarde a instalação dos pacotes finalizar.

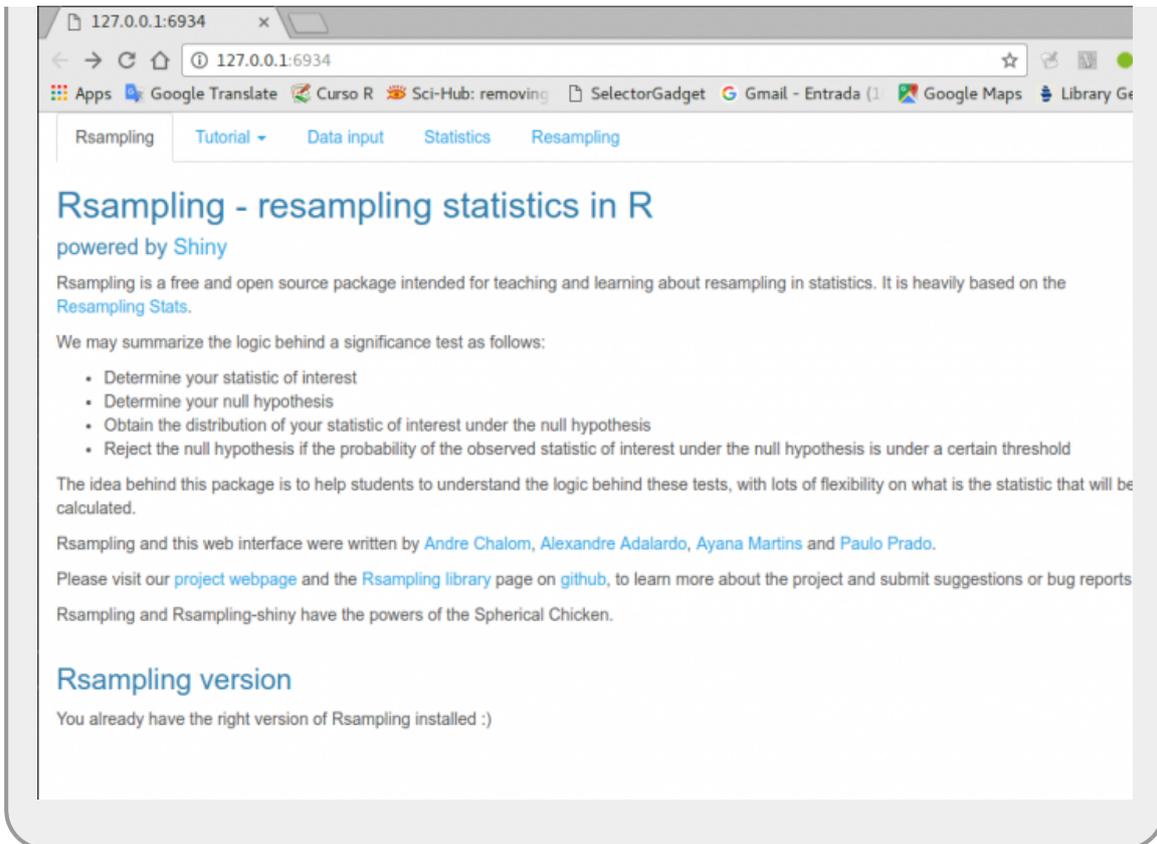
## Abrindo a interface gráfica

Agora digite as seguintes linhas de comando no terminal do R. A primeira linha irá ativar o pacote shiny que acabou de instalar, a segunda irá alterar o idioma para português e a terceira irá abrir a interface Rsampling-shiny no seu navegador, a partir do nome do diretório que descompactou. Sempre que quiser abrir a interface, deve digitar essas três linhas a partir do diretório da disciplina, não há necessidade de instalar os pacotes novamente <sup>3)</sup>.

```
library(shiny)  
language="pt"  
runApp("Rsampling-shiny-1.7.0")
```

Se tudo correu bem, a seguinte interface deverá estar aberta no navegador:

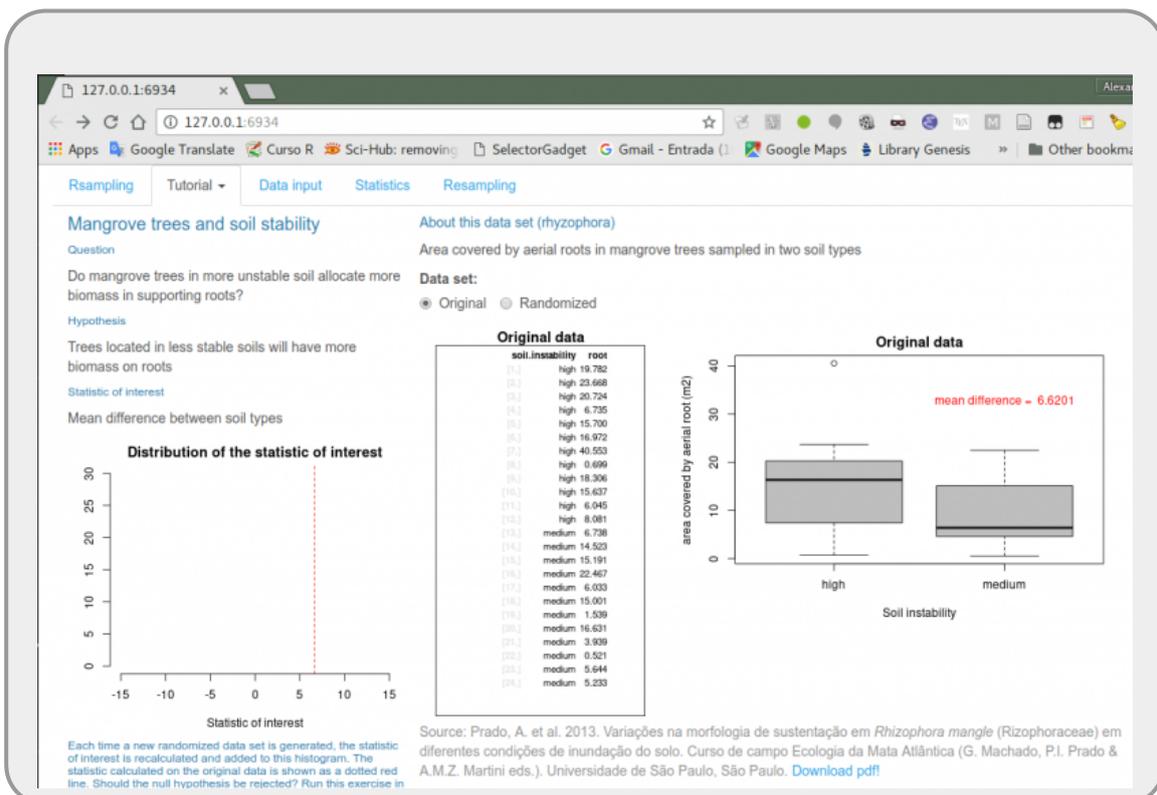




A interface Rsampling tem 5 abas (em português: Rsampling, Tutorial, Entrada de dados, Estatística, Reamostragem). Explore as abas.

## Tutorial Árvores do Mangue

Nesta aula vamos nos ater ao tutorial “Árvores de Mangue”, na aba *Tutorial*.



- não aperte nenhum botão ainda;
- leia a página do tutorial;
- identifique e tente entender a pergunta e a hipótese colocada;
- identifique qual a hipótese nula para a questão colocada;
- formule uma predição esperada, associada à estatística de interesse;
- entenda o que está expresso na tabela central e no gráfico (boxplot) à direita;
- aguarde a discussão desses tópicos com toda a equipe.



### **Coleta de dados**

A descrição da coleta de dados do trabalho original é a seguinte:

“Desenvolvemos o estudo em uma área de manguezal às margens do Rio Una, no município de Peruíbe, litoral sul do estado de São Paulo. Delimitamos duas parcelas de 50 x 50 m, de mesmo relevo, uma próxima à margem do rio, sujeita a um maior tempo de inundação, a qual nos referimos como mais lodosa, e outra no interior do mangue, sujeita a um menor tempo de inundação, a qual nos referimos como menos lodosa. Fizemos uma amostragem arbitrária de 12 indivíduos de *Rhizophora mangle* em cada uma das parcelas. As variáveis utilizadas como medidas de sustentação dos indivíduos foram área de fixação da raiz e número de pontos de apoio de raízes escora no solo. Para medir a área de fixação das raízes de cada indivíduo, aproximamos a área abrangida pelos pontos de apoio das raízes no solo ao formato de uma elipse. Medimos, em metros, o raio maior e o menor dessa elipse. O número de pontos de apoios foi obtido contando todas as vezes que as ramificações das raízes tocavam no solo....”

### **Definindo um cenário nulo**

Entender a hipótese e sua conexão com a estatística de interesse é crucial para compreender o que segue. Faça uma breve pausa e verifique com colegas da turma se entenderam o significado desses conceitos. Ou seja, como a diferença de área de raízes entre os tipos de solo se conecta com a hipótese. Iremos resgatar esses conceitos em outros momentos durante o curso, mas não deixe a aula sem os ter entendido.

Agora precisamos definir um cenário associado à hipótese nula, o nosso cenário nulo. Em geral, o cenário nulo está vinculado à explicação mais simples ou à ausência de efeitos relacionados a nossa hipótese. No exemplo do mangue seria a ausência de efeito na área de raízes relacionada ao tipo de solo.

Imagine que pudéssemos repetir a amostragem, seguindo o mesmo protocolo, muitas vezes. Quais valores da estatística de interesse seriam mais frequentes, partindo da premissa que nosso cenário

nulo fosse verdadeiro?

Será que podemos simular o cenário nulo com nossos próprios dados?

Podemos responder essas questões utilizando a interface do RSampling, como segue abaixo.

### **Usado o RSampling**

- na aba do tutorial, abaixo de **“Conjunto de dados”** selecione **“Aleatorizado”**, depois volte para **“Original”**. Alterne entre essas opções algumas vezes e procure entender o que acontece quando muda a seleção;
-  deixe **“Aleatorizado”** selecionado e clique no botão **“De novo!”**, que fica abaixo do boxplot, algumas vezes;
- compare os resultados dos valores que aparecem no boxplot com os valores que são colocados no gráfico de barras à esquerda;
- clique em **“De novo!”** várias vezes, até que umas das barras chegue ao valor máximo do gráfico à esquerda;
- discuta com os colegas o significado desse gráfico e aguarde a discussão com toda a equipe.

## **Reamostrando os dados**

Caso tenha entendido o tutorial, podemos partir para o teste de hipótese por aleatorização.

- Na aba *Entrada de dados* selecione os dados **“rhyzophora”**;
- Na aba *Estatística* :
  1. em **Estatística**: selecione **“Diferença média entre dois grupos”**;
  2. em **Column 1**<sup>4)</sup>: selecione **“soil.instability”**;
  3. em **Column 2**<sup>5)</sup>: selecione **“root”**;
-  Na aba *Reamostragem* :
  1. em **Tipo de aleatorização**: selecione **“Dentro de colunas”**;
  2. em **Alternativa**: selecione **Maior que**;
  3. em **Número de repetições**: deslize botão para o número 1000;
  4. **Não** selecione **“Com reposição?”**, nem **“Reamostragem estratificada”**;
  5. clique em **Atualizar gráfico**;
- Entenda o que está acontecendo, troque informações com os colegas e aguarde a discussão com os professores

## **Teste Bi ou Unicaudal**

Uma decisão, tomada no começo do tutorial, não foi apresentada anteriormente: o teste feito até agora é unicaudal. O que significa isso? Mude a opção na aba de *Reamostragem* na barra **Alternativa** de **“Maior que”** para **“Bicaudal”**.



- acompanhe o que acontece com o  $p$ -value;
- discorra com os colegas a diferença desses dois testes;
- há mudança na hipótese biológica em questão?

## Exercício

- Caso ainda não tenha feito, anote o seu nome, altura (em centímetros e sem decimais) e sexo (M, F) na planilha [altura alunos](#).
- aguarde todos os alunos preencherem
- salve a planilha no seu computador, usando o formato **.csv**
- importe a planilha para o Rsampling, usando na aba “Entrada de Dados” a opção “Upload file”
- selecione a opção “Cabeçalho?”
- com esse conjunto de dados, teste as seguintes hipóteses usando o Rsampling:
  - altura média entre gêneros é diferente;
  - altura média de estudantes do sexo masculino é maior que de estudantes do sexo feminino.

### **PARA ENTREGAR ANTES DO INÍCIO DA PRÓXIMA AULA**

- 1 - os gráficos finais e o  $p$ -valor para as duas hipóteses
- 2 - a interpretação dos resultados (máximo 5 linhas para cada hipótese)

Obs: tente encaminhar suas respostas pelo formulário abaixo, que estará visível caso esteja logado em uma conta google. Pode ser submetido também pelo [link aqui](#), após login em conta google. Caso encontre problemas na submissão dos formulários, encaminhe o documento à equipe (**[planecousp@gmail.com](mailto:planecousp@gmail.com)**), indicando como “Assunto”: **Teste de Hipóteses**

1)

se não sabe a diferença entre os dois tipos de arquivos, provavelmente irá preferir o zip

2)

Pode haver mais de uma versão do R instalado no computador. Escolha a mais recente em 64b

3)

garanta que a versão do Rsampling-shiny chamada corresponde com a que baixou, no caso abaixo é a 1.7.0

4)

aqui você deve sempre colocar a variável que define os grupos

5)

aqui você deve colocar a variável resposta

From:

<http://labtrop.ib.usp.br/> - **Laboratório de Ecologia de Florestas Tropicais**

Permanent link:

<http://labtrop.ib.usp.br/doku.php?id=cursos:planeco2021:roteiro:02-hipot>



Last update: **2022/02/02 12:00**