

# Carolina Nisa Ramiro



Doutoranda do Programa de Pós-Graduação em Zoologia do Instituto de Biociências da USP. Sob orientação do Prof. Dr. Miguel Trefaut Rodrigues (Laboratório de Herpetologia IBUSP) desenvolvo um projeto de pesquisa que busca caracterizar os compostos químicos que podem ser utilizados como feromônios por espécies brasileiras de lagartos da família Teiidae.

## Meus exercícios

exec

## Trabalho Final

### Plano Principal

#### Problemas com n amostral

Contexto: Pesquisadores, principalmente aqueles que trabalham com vertebrados, frequentemente dependem exclusivamente de espécimes depositados em coleções, limitando assim o tamanho das amostras. Periódicos científicos têm aumentado o rigor quanto à validade real dos resultados obtidos de comparações de amostras de tamanhos muito diferentes, tornando necessária a comprovação estatística de que os resultados apresentados não foram obtidos como consequência do tamanho das amostras.

Proposta: A função irá tratar de n variáveis morfológicas (morfométricas e merísticas) separadas por uma única variável categórica apresentadas em um data frame. Serão geradas reamostragens que igualem o tamanho das duas amostras, aplicando os testes estatísticos para cada uma. Inicialmente será aplicado o teste de equivalência de variâncias nas amostras originais. Caso os dados tenham homogeneidade de variâncias, a função dará continuidade à análise por meio de ANOVA e em caso negativo, Mann-Whitney. As análises estatísticas serão repetidas mil vezes. A função retornará ao usuário uma lista com os valores de p para cada teste realizado, assim como um histograma dos valores de p e a posição do p obtido através da análise dos dados originais.

Parece legal mas como igualar o tamanho das amostras? sorteando quais observações da amostra maior serão removidas? Isso é importante.

Tbm não ficou claro pra mim qual teste será repetido mil vezes: (1) as reamostragens, ou (2) ANOVA/Teste-U, ou os dois. Imagino que sejam só as reamostragens, mas seria legal deixar isso mais explícito.

## Plano Alternativo

### Influência do dimorfismo sexual

Contexto: Trabalhos que comparam populações diferentes da mesma espécie realizam as análises para fêmeas e machos adultos separadamente sem antes realizar um teste de dimorfismo sexual ou sem compreender o real papel do dimorfismo sexual nas diferenças entre as localidades.

Proposta: A função irá tratar de  $n$  variáveis morfológicas (morfométricas e merísticas) e duas variáveis categóricas (sexo e localidade) apresentadas em um data frame. Inicialmente a função irá testar qual o modelo mais adequado para os dados ("sexo", "localidade", "sexo:localidade" ou "sexo, localidade, sexo:localidade") e posteriormente irá aplicar os testes estatísticos (ANOVA ou Mann-Whitney) de acordo com o modelo mais adequado. O resultado será apresentado na forma de um data frame com o sumário do modelo mais adequado e os resultados dos testes aplicados.

Não sei se entendi o propósito dos testes: é testar se o dimorfismo é diferente em cada localidade?

Se sim, apenas o modelo linear com a interação resolve o problema. Mas o modelo linear só funciona se a premissa de normalidade e homocedasticidade dos dados for atendida.

Não seria melhor primeiro fazer um teste de normalidade e homocedasticidade, e depois a partir dele decidir o teste: modelo linear ou teste-U?

Qualquer coisa podemos ir trocando ideia via atualizações nesta página (pretendo visitar aqui noavmente dia 09/06 antes das 12hrs) ou por mensagem. Meu whatsapp é (11) 9-9199-3842

Twister (Matheus Januario)

## Página de Ajuda

amostral

package:nenhum

R Documentation

Repetição de reamostragem de amostra de menor tamanho e aplicação da estatística de interesse (ANOVA ou Mann-Whitney)

Description:

A função reamostra com reposição a categoria com menor  $n$  amostral igualando

seu tamanho com o da categoria de maior n amostral, aplicando então a estatística de interesse (ANOVA ou Mann-Whitney). Esse processo é repetido mil vezes, devolvendo para o usuário um histograma com a distribuição dos valores de p obtidos para cada processo de reamostragem.

#### Usage:

```
amostral<-function(x,cat,cat1,cat2,var,tam,log10,ciclos)
```

#### Arguments:

x dataframe com duas variáveis. A primeira coluna deve corresponder à variável categórica e a segunda à variável a ser testada (morfométrica ou merística), que deve ser numérica.

cat nome da categoria. Deve ser colocado entre aspas para que a função reconheça.

cat1 nome da categoria de menor tamanho amostral. Deve ser colocado entre aspas para que a função reconheça.

cat2 nome da categoria de maior tamanho amostral. Deve ser colocado entre aspas para que a função reconheça.

var nome da variável a ser testada pela estatística de interesse. Deve ser colocado entre aspas para que a função reconheça.

tam tamanho que a amostra de menor n amostral terá após a reamostragem com reposição. Esse valor deve ser igual ao valor do n amostral de cat2.

log10 transforma a variável a ser testada em seu logaritmo decimal para garantir pressupostos de normalidade de distribuição dos valores. Por default, log10=TRUE, caso os valores já estejam em seu logaritmo decimal utilizar llog10=FALSE.

ciclos número de vezes que a função for irá reamostrar os valores e aplicar a estatística de interesse. Por default, ciclos=1000.

#### Details:

Confirmar, antes de executar a função, a classe das variáveis do dataframe e, se necessário, utilizar a função `as.numeric` para a variável a ser testada.

Calcular previamente o valor de p da estatística de interesse para as amostras originais para comparações.

Decimal indicado por ".".

#### Value:

Objeto resulta com os mil valores de p gerados pela estatística de interesse (ANOVA ou Mann-Whitney).

Histograma com a frequência dos valores de p gerados pela estatística de interesse (ANOVA ou Mann-Whitney).

#### Warning:

A função não devolve um único valor de p para que seja feita a tomada de

decisão, mas permite observar a variação nos valores de p igualando o tamanho das amostras e considerando que sua amostral de menor tamanho corresponde à uma amostra confiável da população real de valores para aquela variável.

Ao aplicar Mann-Whitney a função pode devolver warnings devido a um problema intrínseco do teste estatístico quando há repetição de valores dentro da variável testada.

Author(s):

Carolina Nisa  
carolina.nisa@gmail.com

References:

Teste de equivalência de variâncias.  
[https://en.wikipedia.org/wiki/Bartlett%27s\\_test](https://en.wikipedia.org/wiki/Bartlett%27s_test)  
Mann-Whitney-Wilcoxon Test in R.  
<http://www.r-tutor.com/elementary-statistics/non-parametric-methods/mann-whitney-wilcoxon-test>

See Also:

Olhar as funções "bartlett.test", "for", "anova" e "wilcox.test" que calculam valores dentro da função amostral.

Examples:

Dados de altura da cabeça de lagartos em mm (ACA) de duas localidades (norte e sul).

- Para executar ANOVA dentro da função com 150 repetições:

```
ACA<-  
c(10.98,10.84,10.22,11.28,11.20,9.06,9.87,10.64,9.84,10.83,9.47,10.99,10.87,  
10.21,11.11,11.08,9.56,11.61,8.65,10.40,9.08,9.96,9.43,9.69,9.33,9.95,10.12,  
9.73,9.66,11.06,11.15,12.81,11.78,8.45,10.17,8.68,11.09,11.33,10.18,9.13,8.7  
3,10.44,10.23,11.03,9.75,10.95,10.95)  
localidade<-c(rep("norte",16),rep("sul",31))  
dados.anova<-data.frame(localidade,ACA)  
amostral(dados.anova,cat="localidade",cat1="norte",cat2="sul",var="ACA",tam=  
31,log10=TRUE,ciclos=150)
```

- Para executar Mann-Whitney-Wilcoxon dentro da função com 1000 repetições:

```
ACA<-  
c(10.98,10.84,10.22,11.28,10.64,9.84,10.83,9.47,10.99,10.87,10.21,11.11,11.0  
8,9.56,11.61,8.65,10.40,9.08,9.96,9.43,9.69,9.33,9.95,10.12,9.73,9.66,11.06,  
11.15,12.81,11.78,8.45,10.17,8.68,11.09,11.33,10.18,9.13,8.73)  
localidade<-c(rep("norte",11),rep("sul",27))  
dados.anova<-data.frame(localidade,ACA)  
amostral(dados.anova,cat="localidade",cat1="norte",cat2="sul",var="ACA",tam=  
27)
```

# Código da Função

```
#função para lidar com amostras de tamanhos diferentes
#repete a estatística de interesse (ANOVA ou Mann-Whitney)

amostra1<-function(x,cat,cat1,cat2,var,tam,log10=TRUE,ciclos=1000)
#x=dataframe cat=nome da variável categórica cat1=nome da categoria de menor
n cat2=nome da categoria de maior n var=variável a ser testada tam= igual ao
n da cat2 log10=true para transformar dados em log10 ciclos=quantas vezes
repete o ciclo na função for
{
  dados<-x #salva dataframe em objeto dados
  if(log10==TRUE) #transforma os dados em seus log decimais
  {
    dados[,2]<-log10(dados[,2]) #só os dados da variável numérica
  }
  else #os dados já estão em log decimal
  {
    dados<-dados
  }
  variancia<-bartlett.test(x[,2],x[,1]) #cálculo da equivalência de
variância
  p.variancia<-variancia$p.value #salvando só o valor de p do teste de
variância
  dados.cat1<-subset(dados,dados[,1]==cat1) #separando só os dados da
categoria 1
  dados.cat2<-subset(dados,dados[,1]==cat2) #separando só os dados da
categoria 2
  rep.cat1<-rep(cat1,tam) #cria um novo vetor para cat1 com o tamanho que a
amostra vai ter
  resulta<-rep(NA,ciclos) #cria vetor com 1000 NAs que será preenchido na
função for

  #pelos valores de p dos testes

  if(p.variancia>=0.05) #p>=0.05 a função segue realizando ANOVA
  {
    for(i in 1:ciclos) #para repetir ciclos vezes as linhas de comando
    {
      s.cat1<-sample(dados.cat1[,2],tam,replace=TRUE) #aplicando sample para
cat1 e gerando objeto com comprimento igual ao da cat2
      dados.cat1<-data.frame(rep.cat1,s.cat1) #transforma esse sample em um
dataframe
      colnames(dados.cat1)[1]<-cat #nomeando coluna 1
      colnames(dados.cat1)[2]<-var #nomeando coluna 2
      dados<-rbind(dados.cat1,dados.cat2) #construindo dataframe para
ANOVAs.
      lmanova<-lm(dados[,2]~dados[,1],dados) #modelo para realizar anova
      anova<-anova(lmanova) #fazendo anova com as amostras com mesmo tamanho
      resulta[i]<-anova$Pr[1] #preenche objeto resulta com os p gerados na
```

```
ANOVA
}
histograma<-hist(resulta,xlab="p valor
ANOVA",ylab="Frequência")#histograma com as frequências dos p
return(resulta) #retorna para o usuário
}
else #se o p<=0.05 no teste de variância
{
  for(i in 1:ciclos) #para repetir ciclos vezes as linhas de comando
  {
    s.cat1<-sample(dados.cat1[,2],tam,replace=TRUE) #aplicando sample para
cat1 e gerando objeto com comprimento igual ao da cat2
    dados.cat1<-data.frame(rep.cat1,s.cat1) #transforma esse sample em um
dataframe
    colnames(dados.cat1)[1]<-cat #nomeando coluna 1
    colnames(dados.cat1)[2]<-var #nomeando coluna 2
    dados<-rbind(dados.cat1,dados.cat2) #construindo dataframe para Mann-
Whitney-Wilcox
    mann<-wilcox.test(dados[,2]~dados[,1]) #teste de Mann-Whitney-Wilcox
    resulta[i]<-mann$p.value #preenche objeto resulta com os p gerados no
Mann-Whitney
  }
}
histograma<-hist(resulta,xlab="p valor Mann-Whitney",ylab="Frequência")
#histograma com as frequências dos p
return(resulta) #retorna para o usuário
}
```

## Arquivo da Função

[amostral.r](#)

From:

<http://ecor.ib.usp.br/> - **ecoR**

Permanent link:

[http://ecor.ib.usp.br/doku.php?id=05\\_curso\\_antigo:r2017:alunos:trabalho\\_final:carolina.ramiro:start](http://ecor.ib.usp.br/doku.php?id=05_curso_antigo:r2017:alunos:trabalho_final:carolina.ramiro:start)

Last update: **2020/08/12 06:04**

