

Lorena Bueno



Mestranda pelo programa de pós graduação em Ciência Biológicas (Botânica) pela UNESP.

Atualmente eu estudo uma espécie de Leguminosae do cerrado, *Chamaecrista desvauxii*, com o intuito de compreender as interações mutualistas e suas possíveis interferências.

exec

Trabalho Final

Ideia A

Como os dados devem estar no R: O objetivo inicial deve ser um data frame.

Objetivo da função: Calcular média e desvio padrão de grupos dentro de um objeto data frame composto somente por variáveis quantitativas contínuas. A função definirá grupos (conjuntos de linhas) a partir de uma das colunas do data frame que será tratada como coluna referência e será definida em um argumento na função. O número de grupos que será criado também será um argumento da função e definirá os níveis dentro da coluna referência. Os grupos serão formados a partir da coluna referência em ordem crescente.

Sendo assim: A função irá basicamente “agrupar” os seus dados em níveis (que serão definidos por você) a partir de uma coluna referência (também definida por você) dentro do data frame e automatizar o cálculo de média e desvio padrão de cada um dos níveis criados. Essa função irá produzir um novo data frame com as novas colunas criadas a partir da função, que será a coluna dos níveis criados, a coluna de média e de desvio padrão para cada nível. Assim, cada coluna original produzirá duas colunas resultado, uma para a média por nível e outra para o desvio padrão por nível.

Justificativa: Essa seria uma função bastante útil caso você queira categorizar seus dados e calcular estatísticas descritivas a partir de uma coluna específica (com dados contínuos), ao invés de somente calcular a média e desvio de todos os dados (elementos) dentro de cada coluna. Sendo assim, você não precisaria criar outros data frames com apenas aqueles valores separadamente e nem calcular isso de forma mecânica (um por um). Além do mais, você poderia definir e variar o número de níveis (categorias) e recalculá-los instantaneamente, reaplicando a função.

Exemplo baseado em dados fictícios para tornar a função mais explícita: Imaginando que você tenha medidas de altura de várias árvores de uma parcela com 340 indivíduos, e junto tenha outras medidas tomadas no campo, e você queira agrupar os dados e calcular a estatística descritiva dos mesmos em três níveis de altura, e calcular a média e desvio padrão de todas as outras colunas (e.g. diâmetro a altura do peito, índice de área foliar). Sendo assim, essa função irá automatizar a criação dos três níveis, que foi pré-determinado por você, e calcular as médias e desvio padrão dos mesmos níveis. Da mesma forma, você pode mudar de ideia e querer criar mais ou menos níveis depois da primeira rodada, e fará isso rapidamente novamente.

Ideia B

Como os dados devem estar no R: Um data frame onde a primeira coluna se tenha dois ou mais tratamentos, e outras colunas com variáveis contínuas.

Objetivo da função: primeiramente a função vai analisar o dados e determinar se há ou não normalidade nos mesmos, caso sim, a função seguirá para os testes paramétricos, caso os dados não sejam normais, a função irá seguir para os testes não paramétricos.

Justificativa: Automatizar a análise dos seus dados que estarão dispostos em um data frame que tenha uma coluna com diferentes tipos de tratamentos e outras colunas com variáveis contínuas, vai comparar a coluna de tratamentos com cada coluna das variáveis contínuas, e no final devolve uma tabela com os valores do teste estatístico.

Porém, não sei se eu consegui me expressar muito bem ao descrever aqui, acho que a primeira foi onde eu dei o meu melhor pois foi difícil conseguir pensar em duas ideias que fizessem muito sentido e que eu conseguisse me expressar bem.

Olá Lorena. Em primeiro lugar, você poderia incluir as propostas aqui na página, com diagramação e tudo o mais? Pode olhar o código das wikis de alunos dos anos anteriores se estiver com dúvidas de como fazer.

Sobre a proposta A

Está um pouco difícil de entender quais seriam os parâmetros de entrada. Pode fazer uma lista com eles, ou dar um exemplo? Pelo que eu entendi, você receberia como parâmetro uma lista de níveis, por exemplo, níveis=c(0.0,1.0,1.5,3.0) dividiria em três níveis, de 0 a 1, de 1 a 1.5, de 1.5 a 3, é isso que você está pensando?

Achei a proposta muito simples. Sugiro incrementar fazendo gráficos. Você poderia fazer gráficos das outras variáveis separadas pelos níveis. Uma coisa que é interessante de se fazer é fazer gráficos em que a separação em níveis está explícita por cores, incluindo legenda das cores. Por exemplo, se tiver apenas uma outra variável, um histograma, ou se tiver pelo menos duas variáveis (x e y), um gráfico $x \sim y$ onde a cor de cada ponto representa um nível. Outra coisa interessante seria se a função soubesse tratar tanto variáveis contínuas quanto categóricas, fazendo o gráfico que faz mais sentido em cada caso.

Sobre a proposta B

Como você comentou, está um pouco difícil de entender essa proposta. Que tipo de teste você quer fazer, e com que objetivo?

Geral

A proposta B está muito mal delineada, então acho que faz sentido investir na proposta A, mas ela precisa ser mais elaborada.

—*Mali*

Olá, Lorena, vi que você colocou as propostas da função aqui na wiki! Mas ainda falta elaborar melhor a proposta que você vai fazer. Você tem até o dia 9 para colocar mais

detalhes na proposta e receber novos comentários pra te ajudar com a proposta.

—*Mali*

Oi, peço desculpas por não ter respondido antes, mas eu tive um pouco de dificuldades em criar a função, considerando que meu nível no R ainda é iniciante. Eu tentei incorporar um pouco dos seus comentários e sugestões na ideia a. Obrigada pela ajuda.

PAGINA HELP

change.table{unknown} package:unknown R
Documentation

CRIANDO NIVEIS CATEGORICOS

Description:

Calcular média e desvio padrão de grupos dentro de um objeto data frame composto somente por variáveis quantitativas contínuas.

Usage:

```
change.table=function (data, column, group, variable)
```

Arguments:

data: é o data frame, com todas as tabelas e colunas de interesse
column: coluna referência escolhida para criar os níveis de categoria
group: número de níveis de categoria de interesse
variable: variável dependente de interesse para plotar no gráfico

Details:

A função irá basicamente “agrupar” os seus dados em níveis (que serão definidos por você)
a partir de uma coluna referência (também definida por você) dentro do data frame e automatizar o cálculo de média e desvio padrão de cada um dos níveis criados.

Value:

A função retorna uma lista com todos os novos data frames criados a partir do número de níveis de categorias escolhidos.

Author(s):

Lorena Bueno Valadao Mendes
lorena.bvmendes@gmail.com

Botucatu, 25 de junho de 2017

Examples:

```
data=read.table ("example_plant_phenology.csv",header=TRUE, sep=";")  
change1=change.table(data, 6,4,5)
```

CÓDIGO DA FUNÇÃO

```
##FUNÇÃO DISCIPLINA R USP##  
#####  
  
change.table=function (data, column, group, variable)  
{  
  data2 <- data[order (data[,column]),]      ### ordena os dados em ordem  
  crescente pela coluna de interesse ex. coluna 6-> "plant_size"!  
  levels<- cut (data2[,column],breaks=group)   ### cria um vetor categorico  
  com 4 niveis de acordo com a variavel de interesse ex. "plant_size".  
  data3 <- split (data2,levels)                ### divide os dados  
  originais ordenados em quatro conjuntos de dados de acordo com os niveis  
  criados para a variavel de interesse ex."plant_size"  
  m <- data.frame  
  (rbind(sapply(data3[[1]],FUN="mean"),sapply(data3[[2]],FUN="mean"),sapply(da  
  ta3[[3]],FUN="mean"),sapply(data3[[4]],FUN="mean"))) )   ### junta as medias  
  dos 4 grupos  
  s <- data.frame  
  (rbind(sapply(data3[[1]],FUN="sd"),sapply(data3[[2]],FUN="sd"),sapply(data3[[  
  3]],FUN="sd"),sapply(data3[[4]],FUN="sd"))) )   ### junta os desvios  
  padroes dos 4 grupos  
  list ("mean"=m,"sd"=s)  
  par(mfrow=c(1,2))  
  name=as.data.frame(dimnames(m))  
  plot (m[,variable]~m[,column],main="categorized", ylab=name[variable,2],  
  xlab=name[column,2], las=1) ### visualizando como ficou a divisao em quatro  
  grupos (pelas medias)  
  plot(data[,variable]~data[,column], main = "original  
  dataset",ylab=name[variable,2], xlab=name[column,2], las=1) ### visualizando  
  mesmos dados originais.  
  return(data3 <- split (data2,levels)) ###retorna uma lista com todos os data  
  frame criados que você definiu previamente na funcao.  
}
```

##FIM##

Arquivos da função

[funcao_trabalho_final.r](#)

[pagina_help.r](#)

[example_plant_phenology.csv](#)

From:

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

Permanent link:

http://ecor.ib.usp.br/doku.php?id=05_curso_antigo:r2017:alunos:trabalho_final:lorena.bvmendes:start 

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