

Ivan G C Ferreira



Graduado em Ciências Biológicas pelo Instituto de Biociências - USP e Mestrando em Oceanografia Biológica pelo Instituto Oceanográfico - USP.

Na graduação trabalhei com genética e fisiologia de bactérias e agora no mestrado estou me envolvendo mais com a ecologia dos microorganismos, estudando a diversidade e funções de suas comunidades no ambiente marinho.

Exercícios

[Ex1](#) [Ex4](#) [Ex5](#) [Ex7b](#) [Ex8](#)

Propostas

Obs: não consegui decidir um plano como A e um como B, mas tenho as duas idéias.

[Propostas](#)

Trabalho Final

 Trabalho entregue um dia fora do prazo. Será considerado, com ressalva. — [Alexandre Adalardo de Oliveira](#) 2017/06/27 08:12

Arquivos enviados

Arquivos recebidos às 6:30h da manhã do dia 27/06/2017

- [funcaoivanferreira.r](#)
- [excel1mod.csv](#)

Olá Ivan, tudo bem?

Você precisa melhorar um pouco suas propostas para podermos começar a conversar. Primeiro, poste suas ideias diretamente nesta página, não em links, que facilita a correção final e deixa tudo registrado num único lugar.

Segundo, procure apresentar suas ideias com mais clareza e da forma mais detalhada que puder (use e abuse de figuras de exemplo, etc. - sugiro ver como estão organizadas as propostas de alguns colegas de anos anteriores para ter ideia do que estou falando) e deixe bem claro como seria a entrada e a saída da função (com exemplos), quais argumentos pretende criar e o que cada um faria.

Terceiro, coloque uma justificativa para a sua proposta, explicando a relevância que a função teria para você e outros usuários.

Ok? 😊 Qualquer dúvida, estou à disposição!

Abraços! — [Rodolfo Liporoni Dias](#) 2017/06/02 14:25

Comentários Melina

Oi Ivan, como está em cima da hora pra vocês começarem a fazer a função e você ainda não respondeu ao Rodolfo eu resolvi me intrometer...

Primeiro, quando você ler estes comentários, peço que faça o que o Rodolfo pediu, postando as propostas diretamente nesta página (veja exemplos dos outros alunos e do aluno FULANO de Tal). Pois a sua função e a página de Help ficará também nesta página, de preferência logo após os comentários.

Eu também não entendi suas propostas os suficiente pra julgar, mas se fosse pra eu indicar uma para você fazer, seria a proposta B, de coordenadas. Pelo que entendi, você vai ter um `data.frame` com coordenadas que não estão padronizadas para grau decimal (com letras e números misturados) e, após padronizar para grau decimal, calcular as distâncias, é isso? Nesse caso você vai usar funções que trabalham com caracteres (separar o S/N/O/E e identificar se o valor da coordenada será negativo ou positivo) e com funções de coordenadas geográficas (existem pacotes que manipulam este tipo de dado, pode ser útil). Só peço pra tomar cuidado com os caracteres "especiais" (" ' "), pois dependendo da [codificação de caracteres](#) do computador do usuário da sua função, as coisas podem não funcionar.

Achei que a proposta B é mais desafiadora, sugiro tentar fazer esta. Pois a primeira proposta

parece simples e específica demais, já que já existem funções que fazem plots 3D.

— [Melina Leite](#) 2017/06/07 14:51

Olá, visto que não consegui aprimorar as primeiras propostas eu tentei pensar em outras coisas, baseadas na técnica de Bootstrap. De qualquer forma, se julgarem melhor, posso continuar com a proposta que a Melina me respondeu (vou esclarece-la adiante como “Proposta C”).

Proposta A

Função para avaliar a significância de uma sucessão específica de eventos através de Bootstrap. Tirei a idéia de uma análise de sucessão de comunidades ao longo do tempo, em que foi testado por simulação, se a transição de padrão de comunidade “A” para “B” era significativa dentro um conjunto de diversos padrões de comunidade amostrados no mesmo local (Ex: $A \rightarrow B \rightarrow B \rightarrow C \rightarrow B \rightarrow A \rightarrow B \rightarrow \dots$). Na função aqui proposta, qualquer sequência de variáveis categóricas (repetitivas) poderia ser analisada para saber se há algum padrão de sucessão significativo.

Para a entrada da função seria necessário um vetor com variáveis categóricas (ou numéricas a serem transformadas em categóricas).

A função deveria então:

1. Reconhecer ou indexar cada variável categórica;
2. Calcular a frequência de cada transição/sucessão (aqui poderia ter a opção/argumento para reconhecer sucessões com diferentes tamanhos.. 2,3,4,5,6,etc elementos);
3. Realizar a simulação Bootstrap;
4. Analisar a probabilidade de cada sucessão;

A saída da função deve entregar as sucessões mais e/ou menos frequentes que são significativas, conforme uma probabilidade p definida nos argumentos da função (padrão seria 0,05).

Proposta B

Função para analisar agrupamentos em quadrantes de plano cartesiano, por Bootstrap. Supondo, por exemplo, que eu tenha uma área dividida em vários quadrantes, e dentro de cada quadrante uma informação de presença ou ausência. Essa função faria um teste de Bootstrap para analisar se há algum agrupamento significativo dos eventos de presença.

Os dados de entrada seria uma tabela ou data frame com TRUE ou FALSE em cada coordenada xy (talvez seja possível trabalhar com variáveis categóricas na tabela; então a função seria responsável pelo reconhecimento da variável desejada).

A função seria basicamente o Bootstrap, comparando uma estatística dos dados de entrada com o de n simulações com reposição:

1. Calcular para cada célula/quadrante com TRUE qual o número de células vizinhas que também tenham TRUE.
2. Somar esses valores de todas as células e dividir pelo número total de células TRUE.

Obs: Não sei se esta estatística está apropriada pois posso ter problemas com as células da borda do plano.

A função deve entregar a probabilidade de ocorrência dos dados de entrada.

Proposta C

Uma função que crie uma matriz de distância (em unidades métricas) a partir de coordenadas geográficas.

Os dados de entrada seriam duas listas (uma para Latitude e uma para Longitude) com as notações de Latitude ou Longitude (Ex: S 62°06'17.6" OU 62°06'17.6" S) O objetivo é facilitar ao máximo a transformação, então a função deveria transformar os dados em dados numéricos, reconhecendo a letra indicadora de cada coordenada(Ex: o "S" de Sul indicaria que os valores deveriam ser positivos e o "N" de Norte, que seria valores negativos), e assim calcular as distâncias entre os pontos por geometria espacial (eu teria que achar alguma formula genérica que incluísse o raio da Terra.. ficaria um valor aproximado).

— [Alexandre Adalardo de Oliveira](#) 2017/06/14 05:55

Oi Ivan,

Acho que as três novas propostas poderiam ser boas funções. A primeira me parece que está no escopo de modelos de transição de estado e matrizes de transição, a segunda mais relacionada com a densidade da vizinhança e a terceira com transformação de coordenadas. Para vc. não se complicar, sugiro que siga com a da transformação das coordenadas, no caminho que indicou. Alguns pontos para vc. considerar são:

- não há entrada de listas no R: ou as coordenadas entram em colunas de um dataframe ou cada uma entra como um vetor de mesmo tamanho
- a Melina falou sobre o problema de caracteres especiais. É realmente um problema! Para não se complicar com algo que é realmente complicado, minha sugestão é que faça sua função começando com a leitura do arquivo (file) do computador (um csv com duas colunas p. ex) e repasse o argumento `fileEncoding` do `read.table` para sua função, quando fizer a leitura dos dados.
- dê uma olhada nos helps das funções `strsplit` e `grep`. Principalmente dessa última que fala sobre um grupo de funções para manipular string de caracteres
- quando tiver problemas com uma lista (vc. vai ter!), procure a função `unlist` para facilitar sua vida!
- a conversão de coordenadas geográficas para cartesianas não é trivial. Há um modelo (vários) por trás para a conversão da curvatura da terra. Sugiro que faça a conversão desconsiderando essa curvatura (valido para pequenas distâncias, deixe isso claro na documentação) e calcule a distância entre pontos pela diferença entre lat e long em minutos e depois converta os minutos em distância (use o teorema de Pitágoras para converter essas diferenças em distância).
- uma forma tb comum de armazenar coordenadas geográficas é 62°06.36' S, ou seja, decimais de graus. Coloque um argumento na sua função para que o usuário indique se os dados estão no formato de decimais em graus ou em minutos e trate os dois tipos.

BOM TRABALHO!

Função

Olá, Não devo conseguir postar hoje (25/07) a função. Tive problemas principalmente por trabalhar no Linux com arquivos gerados no Windows. Não consegui me resolver com o fileEncoding como sugerido. Estou conseguindo trabalhar a função com arquivos gerados também no Linux e, se possível, depois eu tento trabalhar a questão do encoding.

Ivan Ferreira.

— *Alexandre Adalardo de Oliveira* 2017/06/25 19:08 Vc precisa postar algo para podermos avaliar ou nao teremos como dar a sua nota. A sua justificativa não é aceitável. Poste o que conseguiu fazer ainda hj, amanhã o wiki estará fechado para edição.

From:

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

Permanent link:

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

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