

Jana Menegassi del Favero



Doutoranda em Oceanografia Biológica com tese intitulada: Flutuações inter-decadais (1970-2010) na distribuição e abundância de larvas e ovos de anchoita (*Engraulis anchoita*) na Plataforma Continental Sudeste do Brasil.

exec

PROPOSTAS

Principal

Meus dados são: Cp (comprimento padrão em mm), estações do ano (primavera, verão, outono, inverno), espécies de peixes (no caso amostradas na zona de surfe da Ilha Comprida e Ilha do Cardoso). Sabe-se que juvenis de peixes recrutam na zona de surfe de praias e nela permanecem por um curto período de tempo. Fazendo 4 histogramas (um por estação) por espécie é possível notar, em alguns casos, o deslocamento da moda de comprimento padrão ao longo das estações, permitindo afirmar se a espécie está utilizando a área para o crescimento e alimentação ou se ela simplesmente usa a área como rota de migração (no caso entre o mar e o estuário). Assim sendo, o objetivo da proposta seria criar uma função que elaborasse um único gráfico com 4 curvas de densidade (uma por estação) e testasse se a curva se deslocou de forma crescente ao longo do eixo X (ou seja, se a classe modal de CP populacional tem aumentado ao longo do ano desde o recrutamento da espécie na área). Não sei ao certo ainda que teste usar, talvez o teste não paramétrico de Friedman seria uma opção.

Plano B

Durante quase 2 anos trabalhei na elaboração da Coleção Biológica Marinha do Instituto Oceanográfico (ColBIO) e de lá vou retirar meus dados para a realização do meu projeto de doutorado que será com flutuações de longo prazo em larvas e ovos de anchoita. Uma primeira dificuldade encontrada é como ter uma visão geral dos dados disponíveis. (Exemplo: quantidade de projetos, o projeto X foi realizado em tal data, em quais estações hidrográficas, possui tais dados disponíveis e x amostras). Só para sintonizar, atualmente o banco de dados está sendo montado em planilhas do Excel padronizadas e possuem 189 colunas e algumas planilhas chegam até 40.000 linhas! Assim sendo, o objetivo da proposta B seria criar uma função que resumisse as informações contidas nas planilhas de forma simples e de fácil visibilidade, tanto uma visão geral da planilha, tanto escolhendo se por projeto, por espécie, por ano, etc. Um argumento da função seria a elaboração de um mapa com as estações hidrográficas plotadas. Os resultados poderiam sair em formato data.frame, sendo as variáveis que são nomes (caracteres) seriam somadas (ex. X projetos, X espécies) e as variáveis abióticas a média.

Ficou confuso né?! As duas propostas me ajudariam muito!!!!

PÁGINA DE AJUDA

```
plot.test.ecdf                                     package:unknown  
R Documentation
```

```
**Plot multiple ecdf and tests its differences**
```

```
**Description**
```

Constrói em um único gráfico diferentes funções de distribuição empírica acumulada e testa, através de testes de permutação, se há diferenças crescentes nos valores utilizados para formarem as funções.

```
**Usage**
```

```
plot.test.ecdf(x,y)
```

```
**Arguments**
```

x: vetor numérico de valores de dados observados.

y: vetor de fatores do mesmo comprimento de x.

```
**Details**
```

O teste de permutação será realizado par a par, e para tanto é preciso que y seja um fator com os níveis de importância definidos pelo usuário. A função testará o nível 1 com o nível 2, o nível 2 com o nível 3 e assim sucessivamente.

```
**Value**
```

Um gráfico múltiplo de distribuição empírica acumulada e os valores de p para os diferentes pares testados.

```
**Author(s)**
```

Jana Menegassi del Favero

janamdf@usp.br

```
**See Also**
```

```
ecdf, plot, sample
```

```
**Examples**
```

```
#Leia a planilha de dados  
teste <- read.csv("teste.csv", sep=";")  
# Transforme y em fator e de níveis de importância  
teste$estacao <- factor(teste$estacao,  
levels=c("Summer", "Autumn", "Winter", "Spring"))  
plot.test.ecdf(teste$CP, teste$estacao)
```

CÓDIGO DA FUNÇÃO

```
plot.test.ecdf<-function(x,y)
{
  #parte grafica#
  l<-tapply(x,y,list)
  par(mfrow=c(1,1))
  grafico<-plot(ecdf(unlist(l[1])), pch=1, main="")
  for(i in 2:(length(l)))
  {
    lines(ecdf(unlist(l[i])), pch=i)
  }
  legend("right",c(names(l)), pch=1:i)
  #parte teste permutacao pareada#
  p<-vector("list",nlevels(y)-1)
  result<-vector("list",nlevels(y)-1)
  s<-vector("list",nlevels(y)-1)
  grupo<-vector("list",nlevels(y)-1)
  for(j in 1:(nlevels(y)-1))
  {
    result[[j]]<- (rep(NA,1000))
    result[[j]][1]<-diff(tapply(x,y,mean)[])[j]
  }
  for(k in 1:(nlevels(y)-1))
  {
    s[[k]]<-sample(c(x[y==levels(y)[k]],x[y==levels(y)[k+1]]))
    grupo[[k]]<-c(y[y==levels(y)[k]],y[y==levels(y)[k+1]])
    for(i in 2:1000)
    {
      s1<-sample(s[[k]])
      g1<-grupo[[k]]
      result[[k]][i]<-diff(tapply(s1,g1,mean)[])
    }
    p[[k]]<-(sum(result[[k]]>=result[[k]][1])/length(result[[k]]))
  }
  return(p)
}
```

ARQUIVOS

[plottestecdf.r](#) [teste.csv](#)

Comentários

PI

A primeira está clara e é viável. Tente pensar não só nos seus dados e fazer uma função apenas para eles: busque uma função genérica que ajude outras pessoas que trabalhem com o assunto. Defina claramente entradas e saídas. E como tempo é curto, vc pode dividir o trabalho em duas etapas: 1) Resultados gráficos, 2) teste estatístico. Sobre o teste, considere incluir um teste por permutação ao invés de um não-paramétrico como vc propôs. Sobre o gráfico, funções de probabilidade/densidade acumulada são mais eficientes para evidenciar diferenças em distribuições do as funções de densidade, veja ecdf.

A segunda de fato ainda não está clara, principalmente pq é possível fazer muitos resumos diferentes. Especifique um deles. De maneira geral, antes de criar uma função é essencial definir sua entrada e saída (campos 'arguments' e 'value' da página de ajuda).

Resposta

Oi Paulo, como você mesmo disse, o tempo é curto, então vou focar mesmo na primeira! Obrigada

Chalom

Jana, a sua proposta A eh ótima. A B me parece mais um projeto de banco de dados do que de R (e pelamordedeus verifiquem a possibilidade de usar um banco relacional, como MySQL, ao invés de Excel, para guardar esses dados. Da um certo trabalho pra aprender e implementar, mas os beneficios são INFINITOS!)

Resposta

Oi Chalom, faz um tempo que estou tentando trabalhar na proposta A, já desencanei da B há algum tempo. Mas estou tendo muitos problemas... Aquele dia que nos encontramos você disse para eu escrever email... qual teu email? Porque escrever pelo "send email" do Forum não está rolando... Obrigada 😊

From:

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

Permanent link:

http://ecor.ib.usp.br/doku.php?id=05_curso_antigo:r2013:alunos:trabalho_final:janamdf:start 

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