

# Nancy Kazumi Taniguchi



Bacharel em Oceanografia, Mestranda em Oceanografia Geológica, Instituto Oceanográfico da Universidade de São Paulo.

Laboratório de Bioindicadores Ambientais

palavras chaves: foraminíferos planctônicos, paleoceanografia, paleoclimatologia, Atlântico, Foz do Parnaíba, Foz do Amazonas, holoceno, isótopos, TSM.

## Exercícios

exec

## Minha Proposta

Os foraminíferos são protistas eucariontes e constituem o grupo mais diverso de organismos microscópicos com conchas carbonáticas nos oceanos atuais. São amplamente aplicados em estudos ecológicos de caracterização microfaciológica e monitoramento ambiental, desde regiões costeiras até áreas oceânicas profundas. São considerados bons bioindicadores, pois são abundantes em sedimentos, apresentam alta diversidade de espécies, ciclo de vida relativamente curto, ampla distribuição geográfica, além de serem sensíveis às variações químicas e físicas do ambiente. Assim, estudos de caracterização e monitoramento constituem parcialmente o diagnóstico ambiental dos meios físicos, bióticos e socioeconômicos que faz parte do documento técnico-científico denominado Estudo de Impacto Ambiental (EIA). A confecção e elaboração deste documento está inserido no processo de licenciamento ambiental de obrigação legal prévia à instalação de qualquer empreendimento ou atividade potencialmente poluidora ou degradadora do meio ambiente.

### Proposta A

Esta proposta consiste na automação da análise estatística para atividades de caracterização microfaciológica de bacias sedimentares no qual terão como entrada de dados planilhas do tipo texto, contendo dados referente à identificação e posição das amostra (identificação da amostra e réplicas), densidade de espécies de foraminíferos, parâmetros bióticos, dados sedimentológicos como %lama, %areia, diâmetro médio e dados geoquímicos como concentração de carbonato (%CaCO<sub>3</sub>), carbono orgânico total (COT) e nitrogênio total (NT). Esta função realizará a leitura e o rearranjo dos dados para execução de funções estatísticas como análise de agrupamento e análise de componentes principais e terão como saída gráficos, figuras e resumos dos parâmetros estatísticos encontrados. Os dados referem-se à espécies de foraminíferos vivos encontrados em sedimento de superfície em uma determinada área.

### Proposta B

Esta proposta consiste na automação da análise estatística porém num contexto paleoceanográfico. Dados paleoceanográficos são relacionados a sedimentos provenientes de testemunhos onde sedimentos de topo representam associações de espécies do período mais recente e a base, condições passadas. Difere-se da caracterização citada na proposta A por tratar de organismos preservados de uma determinada posição, não possuem réplicas e podem refletir mudanças de condições ambientais/oceanográficas ao longo do tempo. Os parâmetros de entrada serão planilhas do tipo texto contendo informação das amostras em profundidade com parâmetros bióticos, sedimentológicos e geoquímicos mencionados na proposta A. Serão obtidos como saída gráficos, figuras e resumos dos parâmetros estatísticos de análise de agrupamento e análise de componentes principais.

## Comentários

A proposta é muito específica e tem cara de um scrip de análise de dados e não de uma função. Pense em trabalhar o problema de maneira mais geral que funcione com conjuntos de dados parecidos. No Task View de Multivariate Statistics e no Environmetrics você encontra caminhos para os pacotes que já fazem coisas parecidas e podem te ajudar a estruturar a proposta com uma cara de função. Se você pensar em construir uma função para fazer análises estatísticas note que as duas propostas são iguais! afinal fazem a mesma coisa. Pouco importa para a função se são bichos verdes ou vermelhos, mas sim o tipo de dado que você insere e o tipo de saída que ela retorna. Precisa reestruturar as propostas e pensar em um plano B que seja de fato diferente do A.

Sara Mortara

## Justificativas

Olá Sara, obrigada pelo comentário... vou seguir suas dicas e refazer as propostas.

Bem... lendo as propostas apresentadas, realmente elas são muito parecidas, porém, acho que preciso esclarecer alguns pontos. Essas propostas foram elaboradas em função das atividades realizadas no Laboratório de Biodindicadores Ambientais do Instituto Oceanográfico. Atualmente as análises citadas são realizadas em diversos softwares dentre eles, parte é feito no software PAST, outro no MVSP e outro no STATISTIC. Elaborei estas propostas com o objetivo de substituir estes três aplicativos ou mais através do uso do software R considerando também o fato de que as análises citadas são realizadas na maioria dos projetos executados no laboratório.

A proposta A, está relacionada a caracterização, numa visão espacial onde a intenção é verificar "padrões" como por exemplo, dada uma bacia sedimentar, se as comunidades são diferentes em função das profundidades, se são "setorizados", como região norte pode ser diferente da região sul, se cada transecto apresenta coisas distintas ou se são todos iguais, se a distribuição das espécies está relacionado com profundidade e presença nutrientes, etc. Os dados de entrada, são dados de superfície de uma área e com réplicas. Futuramente, caso haja projetos de monitoramento nestas bacias sedimentares a partir desta mesma planilha poderia se comparar tanto no tempo ou mesmo quantificar as diferenças/ semelhanças com outras bacias sedimentares.

A proposta B, está relacionado com paleoceanografia. Os parâmetros de entrada são espécies, dados de sedimentologia e geoquímica mas num contexto pontual e comparações no tempo geológico., ou seja, cada profundidade do testemunho é relacionado a uma idade geológica e que vão ser

comparadas entre elas. Neste caso não há replicas e o que se busca nestas análises é comparar semelhanças ou diferenças ao longo do tempo e relacioná-las as mudanças dos processos atuantes de um determinado local/posição. Por exemplo, semelhanças ou diferenças entre periodos glaciais e interglaciais, mudanças em função alterações das condições hidrodinamicas desta posição ou mesmo em função da variação do nivel relativo do mar entre outras coisas. As comparações são todas de um mesmo local mas ao longo do tempo.

Uma outra proposta em que eu pensei foi desenvolver uma função para aplicação de uma metodologia denominada Técnica de Análogos Modernos. Esta técnica tem como objetivo inferir a temperatura de superfície do mar (TSM) a partir de uma associações de espécies conhecida, ou seja, os dados de associações de espécies de foraminiferos planctonicos de uma amostra são comparados a associações de espécies de foraminiferos de um banco de dados global onde um índice de similaridade é aplicado para cada associação do banco de dados. Cada associação do banco de dados estará associada a uma TSM e aquela que apresentar maior similaridade, será sugerido a TSM correspondente. Os parâmetros de entrada seria as associações de espécie de uma amostra e uma outra planilha que poderia ser interna da função ou externa contendo associações de espécies e TSM referencia. Como saída, a função apresentaria a TSM cuja associações teve menor valor de dissimilaridade.

O que você acha destas colocações? Esta terceira proposta, seria uma função mais adequada no contexto da disciplina? 😊

## Comentario

Oi Nancy,

Acho que a proposta A, apesar dos pontos colocados pela Sara, pode ser bastante util, pelo que vc. diz da rotina do seu laboratorio. Vc. precisa ser mais especifica na apresentacao e menos especifica na utilização: “funções estatística como análise de agrupamento e análise de componentes principais e terão como saída gráficos” e muito vago. Faça apenas uma coisa de cada vez e defina bem o que sera, por exemplo: implemente um indice de similaridade entre os objetos (pontos de amostra) e retorna a matriz de similaridade e um cluster das localidades baseado nessa matriz. Isso ja e um bom desafio se vc. partir do zero sem usar nenhum pacote. Alem disso, pode ser usado por qq um que queira agrupar objetos (parcelas, pontos de coletas, individuos, especies...) usando esse indice. Caso termine essa parte, implemente uma segunda parte, p.ex: possibilidade de escolher o tipo de indice (duas opções), gerar um intervalo de confiança para as distancias... Ajuste a proposta nesse sentido para que possamos depois avaliar o resultado pelo que vc. se compromete a fazer. — [Alexandre Adalardo de Oliveira](#) 2013/03/25 11:40

## Proposta de Trabalho Final

Elaborar uma função que apresente como resultado uma matriz de distâncias calculadas a partir de uma planilha de dados texto nos quais linhas referem-se às amostras e colunas representem espécies ou outros fatores. O resultado apresentará uma matriz de distâncias calculadas entre todas as amostras (linhas). Como entrada da função deverão ser indicados se a tabela de dados deverá ser tratado com valores absolutos ou valores relativos (frequencia relativa à amostra) e o método a ser aplicado para o cálculo da distância tendo como opções os métodos as distâncias Euclidiana, Bray-

Curtis, Manhattan, Jaccard ou Índice de Morisita). Além da matriz de distância, a função apresenta um cluster baseado na matriz de resultados.

## Página de Ajuda / Help - função distância

distancia

package:unknown

R Documentation

Esta função calcula e retorna uma matriz de distâncias entre as linhas de uma matriz de dados.

### Description:

Esta função calcula e retorna uma matriz de distâncias entre objetos. A matriz de dados deve conter nas linhas os objetos a serem comparados e nas colunas, as atribuições de cada objeto. As distâncias serão calculadas a cada par de objetos (linhas) e apresentadas na forma de matriz.

### Usage:

```
distancia(x, freq="ABS", method="BRAY-CURTIS")
distancia(x, freq="ABS", method="EUCLIDEAN")
distancia(x, freq="ABS", method="JACCARD")
distancia(x, freq="ABS", method="MORISITA")
distancia(x, freq="ABS", method="MANHATTAN")
```

### Arguments:

`x` matriz de dados onde linha é objeto a ser comparado e colunas as atribuições de cada objeto.  
`freq` valor lógico que indica se matriz de dado deve ser convertido em frequência relativa ou manter em valores absolutos  
`method` indoca método a ser aplicado para o calculo das distâncias. As opções disponiveis são: BRAY-CURTIS, EUCLIDEAN, JACCARD, MORISITA, MANHATTAN.

### Details:

As linhas da matriz de dado deve conter os objetos a serem comparados, por exemplo, as amostras. As colunas conterão as atribuições do objeto, por exemplo, espécies.  
A definição dos parâmetros `freq` e `method` são de preenchimento obrigatório. O método de calculo da distância a ser aplicado deve estar escrito em

maiúsculo conforme os exemplos.

Valores NA presentes na matriz de dados serão substituídos por zero.

freq="ABS": valores da matriz de dado serão mantidos.

freq="RELATIVE": valores da matriz de dados original serão convertidos em valores relativos. Os valores de cada colunas serão divididos pela somatória de cada linha.

Method:

EUCLIDEAN: corresponde a distância geométrica entre dois objetos no plano. Corresponde a distância em linha reta entre dois pontos obtida pelo comprimento da hipotenusa

de um triângulo retângulo onde dados dois pontos (X1, Y1) e (X2, Y2) a distância é calculada como  $\sqrt{(Y1-Y2)^2+(X1^2)-X2}$ .

Distancia pode variar de 0 a infinito

BRAY-CURTIS: Baseado nas distâncias calculadas a partir do módulo das diferenças de densidades entre as amostra para uma dada espécie.

Distancia varia de 0 (similar) a 1 (dissimilar).

$B = \frac{\sum(X_{ij}-X_{jk})}{\sum(X_{ij}+X_{jk})}$ , onde B = distância calculada, X<sub>ij</sub> e X<sub>jk</sub> = número de indivíduos da espécie i nas amostra j e k, sum= somatória

MANHATTAN (ou city-block): distância baseada na soma das diferenças absolutas das variáveis

$M = \sum(\text{abs}(X_{ij}-X_{ik}))$ ; M = distância, X=objeto i, Y= objeto j

MORISITA: índice proposto por Morisita (1959) para medir a similaridade entre duas comunidades. É interpretado como probabilidade de que um indivíduo retirado

da amostra j e um indivíduo retirado da amostra k pertençam à mesma espécie, dividido pela probabilidade de que dois indivíduos retirados das amostra j e k pertençam à mesma espécie. Varia entre 0 a 1, empregado em dados de abundância

JACCARD: definido pelo tamanho da interseção dividido pelo tamanho da união dos itens de dois conjuntos. Baseia em dados de presença e ausência onde 1 é presença e 0, ausência.

Valores da matriz de dados são convertidos em matriz de 0 e 1, onde 1 é assumido para todos os valores maior que 0 da matriz original.

Distância de Jaccard Obtido por:

$$J = \frac{c}{a+b-c}$$

Onde: J = Coeficiente de Jaccard a= numero de atributos do objeto 1 b=

numero de atributos do objeto b, c= numero de atributos comuns a ambas objetos.

Value:

A função retorna uma matriz de distâncias calculadas entre objetos (linhas)

x: é uma tabela de dados com classe data.frame

freq: "ABS" ou "RELATIVE".

method: "BRAY-CURTIS", "EUCLIDEAN", "JACCARD", "MORISITA", "MANHATTAN"

Warning:

Valores NA presentes na matriz de dados serão substituídos por zero.

freq e Method são de preenchimento obrigatório e devem estar entre "aspas" e MAIUSCULA, conforme exemplo: `distancia(x,freq="ABS",method="BRAY-CURTIS")`

Author(s):

Nancy Kazumi Taniguchi  
nancykazumi@usp.br

References:

Bray, J. R. & CURTIS, J. T. 1957. An ordination of the upland forest communities of Southern Wisconsin. *Ecological Monographs*, 27: 325–349.

Hair, Black, Babin, Anderson & Tatham – Análise multivariada de dados (6a edição) – Bookman (2006).

Krebs, C.J.; 2001. *Ecology: the experimental analysis of distribution and abundance*. Benjamin Cummings, an imprint of Addison Wesley Longman, Inc.. San Francisco, California.

Krebs, C.J. 2013. *Ecological Methodology*, 3rd ed. (in prep) Chapters revised to date (14 March 2013) are available to download for evaluation and review (PDF files) <http://www.zoology.ubc.ca/>

Pinto-Coelho, R.M.; 2000. *Fundamentos em ecologia*. Artmed Editora. Porto AlegreRS, 252p.

Zar, J.H.; 1999. Biostatistical analysis. 4ª Edição. Prentice-Hall, New Jersey.

Examples:

```
#construindo uma matriz de dados
especies <-
c("especie1","especie2","especie3","especie4","especie5","especie6","especie
7","especie8","especie9","especie10")
amostra1 <- rpois(10,5)
amostra2 <- rpois(10,5)
amostra3 <- rpois(10,5)
amostra4 <- rpois(10,5)
amostra5 <- rpois(10,5)

dado <- data.frame(amostra1,amostra2,amostra3,amostra4,amostra5)
dado <- t(dado)
colnames(dado) <- paste(especies)
rownames(dado) <-
paste(c("amostra1","amostra2","amostra3","amostra4","amostra5"))
dado

#executando a função para planilha dado construída
distancia(dado,freq="ABS",method="BRAY-CURTIS")
distancia(dado,freq="RELATIVE",method="BRAY-CURTIS")
distancia(dado,freq="ABS",method="EUCLIDEAN")
distancia(dado,freq="RELATIVE",method="EUCLIDEAN")

#teste com arquivo csv e valores NA presentes na matriz de dados.
# primeiramente faça o download do arquivo Planilha_teste.csv disponível no
link junto com a função
teste <- read.table("Planilha_teste.csv", row.names=1, header=T, sep=";",
dec=".", as.is=T)

#executando a função para planilha teste
distancia(teste,freq="ABS",method="BRAY-CURTIS")
distancia(teste,freq="RELATIVE",method="BRAY-CURTIS")
distancia(teste,freq="ABS",method="EUCLIDEAN")
distancia(teste,freq="ABS",method="JACCARD")
distancia(teste,freq="ABS",method="MORISITA")
distancia(teste,freq="ABS",method="MANHATTAN")
```

## Código da função distancia

```
distancia <- function(dado,freq="ABS",method="BRAY-CURTIS")
{
```

```
x <- dado
n.amostras<-dim(x)[1]
n.colunas<- dim(x)[2]
x[is.na(x)]<-0 #tudo o que for NA é substituído por 0.
##Convertendo a tabela de dados para Freqüencia absoluta ou freqüencia
relativa
#padrão é freqüencia absoluta
if (freq=="RELATIVE")
{
  xtemp <- x
  xtemp$temp_total <- apply(xtemp,MARGIN = 1, FUN = sum)
  temp.relative <- xtemp/xtemp$temp_total
  x <- temp.relative[,1:n.colunas]
}
###Calculando distancia entre todas as amostras
#criando matriz de resultados
result.dist <- matrix(data=NA,nrow=n.amostras,ncol=n.amostras)
colnames(result.dist) <- paste(rownames(dado))
rownames(result.dist) <- paste(rownames(dado))
#Criando matriz de resultados para o método Jaccard
result.jacc <- matrix(data=NA,nrow=n.amostras,ncol=n.amostras)
colnames(result.jacc) <- paste(rownames(dado))
rownames(result.dist) <- paste(rownames(dado))
#começando a calcular as distancias entre as amostras - Distancia
Euclidiana
i=1
j=1
#Métodos disponíveis
#method = "BRAY-CURTIS"
#method = "JACCARD"
#method = "EUCLIDEAN"
#method = "MORISITA"
for (i in 1:n.amostras) #i= contador de linha/amostras
{
  j=i+1 #j = contador de amostra+1 (comparar com as
demais amostras)
  for (j in 1:n.amostras)
  {
    d1 <- x[i,] #d1 pega os valores da linha i - amostra A
    d2 <- x[j,] #d2 pega os valores da linha j - amostra B
    if (method == "EUCLIDEAN")
    {
      result.dist[i,j] <- sqrt(sum((d2-d1)^2))
    }
    if (method == "BRAY-CURTIS")
    {
      result.dist[i,j] <- sum(abs(d1-d2))/sum(abs(d1+d2))
    }
    if (method == "MANHATTAN")
    {
```



```

    result.dist[i,j] <- sum(abs(d1-d2))
  }
  if (method == "MORISITA")
  {
    n1 <- sum(d1) # total number of individual in sample i
    n2 <- sum(d2) # total number of individual in sample j
    lambda1 <- (sum(d1*(d1-1)))/(n1*(n1-1))
    lambda2 <- (sum(d2*(d2-1)))/(n2*(n2-1))
    result.dist[i,j] <- (2*sum(d1*d2))/((lambda1+lambda2)*n1*n2)
  }
  if (method == "JACCARD")
  {
    x.jacc <- x
    x.jacc[x.jacc>0]=1 #substitui por 1 todos os valores da tabela
que for maior que 0
    d1 <- x.jacc[i,]
    d2 <- x.jacc[j,]
    comum<- sum(d1>0 & d2>0)
    result.jacc[i,j] <- comum/(sum(d1)+sum(d2)-comum)
    result.dist <- result.jacc
  }
}
j=i
}

display <- round(result.dist, digits = 3) #formatar para 3 casas
decimais os resultados
#plotar cluster
fig <- as.dist(result.dist)
x11()
plot(hclust(fig),hang = 0.1, main = method, sub = "amostras", xlab = "
", ylab = "distância")
return(display)
}

```

## Arquivos para download - função distância

Caso haja necessidade, você pode fazer o download da função distância, o help , arquivo para teste da função e arquivo Planilha\_teste.csv ou solicitar esclarecimentos pelo email [nancykazumi@usp.br](mailto:nancykazumi@usp.br)

[funcao\\_distancia.r](#)

[help\\_distancia.txt](#)

[teste\\_funcao\\_distancia.r](#)

# Proposta de Trabalho Final - Proposta C

Elaborar uma função para calcular valores de anomalia de dados que apresentem variações periódicas ao longo do tempo, por exemplo, temperatura da superfície do mar (TSM). Esta função deverá ser aplicado para uma matriz de dados com 3 dimensões (x,y,z) onde x e y são parâmetros espaciais e z, parâmetro temporal. Para cada série temporal, ou seja, em cada posição (x,y) será ajustada uma curva por regressão linear e determinados seus coeficientes pelo método dos mínimos quadrados conforme a equação:



onde  $\hat{Y}$  representa a variável em questão modelada,  $t_i$  é o tempo da série temporal e os coeficientes  $b$  são determinados através do método de mínimos quadrados.  $b_1$  e  $b_2$  referem-se aos coeficientes da tendência linear,  $b_3$  e  $b_4$  são coeficientes para o ajuste anual e  $b_5$  e  $b_6$  do ciclo semi-anual,  $T$  é igual ao período, por exemplo 12 meses e  $t_i$  o número de meses da série temporal.

Os valores de anomalia serão obtidos através da equação:  $Y' = Y - \hat{Y}$ , onde  $Y'$  é a anomalia,  $Y$  os valores originais da série temporal e  $\hat{Y}$  valores do modelo ajustado com ciclo anual e semi-anual.

Esta metodologia é muito empregada para obtenção de valores de anomalia de TSM nos estudos relacionados à oceanografia e foi desenvolvida baseada no conceito de regressão multivariada aplicada em matrizes para um modelo de estimativa linear.

A referência base desta função é: Emery, W. J and Thomson, R. E.. 2001. *Data analysis methods in physical oceanography*. Elsevier Science.

## Página de Ajuda / Help - função MMQ

mmq\_function

package:unknown

R Documentation

Esta função calcula e retorna valores de anomalias obtidas a partir da diferença entre uma série temporal de um parâmetro qualquer e uma série temporal ajustada

por regressão linear onde os coeficientes de ajustes são determinados pelo método dos mínimos quadrados aplicados à uma matriz de 3 dimensões.

### Description:

Esta função calcula e retorna uma matriz de anomalias obtidas a partir de uma matriz com 3 dimensões (x,y,z), onde x e y são parâmetros espaciais e z é o

parâmetro temporal. Esta função ajustará por regressão linear uma curva com períodos anual e semi-anual para cada posição (x,y) e seus coeficientes serão obtidos

através do método de mínimos quadrados aplicados em matrizes. A anomalia

será obtida a partir da diferença entre o valor do dado original e o valor da curva ajustada

para cada posição. Veja o tópico "Multivariate regression" em Emery, W. J and Thomson, R. E.. 2001. Data analysis methods in physical oceanography. Elsevier Science.

Usage:

```
mmq_function(X,T=12,file_name="anomalia_result.RData")
```

Arguments:

X matriz de dados com 3 dimensões onde x,y são parâmetros espaciais e z, temporal.

T refere-se ao período dos ciclos.

rmNA (rmNA==TRUE ou rmNA==FALSE). Se TRUE omitirá os valores NA's existentes na matriz de dado original. Se FALSE os NA's existente não serão omitidos.

file\_name nome do arquivo em que serão salvos as anomalias calculadas.

Details:

Esta função é aplicada para matrizes com 3 dimensões (x,y,z). (x,y) são parâmetros espaciais, por exemplo, longitude e latitude, e z é o tempo. Ou seja,

podemos trabalhar com dados por exemplo de temperatura de superfície do mar (TSM) obtidas de uma determinada região ao longo de um período. Para cada posição (x,y)

será obtida através de regressão linear uma curva modelo com ciclos anual e semi-anual. O valor da anomalia será dado em função da diferença entre valor original

e valor da curva ajustada.

$$\hat{Y}_i = b_1 + b_2 t_i +$$

$$b_3 \sin((2\pi/T)*t_i) + b_4 \cos((2\pi/T)*t_i) + b_5 \sin((4\pi/T)*t_i) + b_6 \cos((4\pi/T)*t_i)$$

onde:

$\hat{Y}$  representa a variável em questão modelada,

$t_i$  é o tempo da série temporal,

coeficientes b são determinados através do método de mínimos quadrados.

$b_1$  e  $b_2$  referem-se aos coeficientes da tendência linear,

$b_3$  e  $b_4$  são coeficientes para o ajuste anual,

$b_5$  e  $b_6$  do ciclo semi-anual,

T é igual ao período, por exemplo 12 meses e

$t_i$  o número de meses da série temporal.

Os valores de anomalia serão obtidos através da equação:

$$Y' = Y - \hat{Y},$$

onde  $Y'$  é a anomalia,  
 $Y$  os valores originais da série temporal e  
 $\hat{Y}$  valores do modelo ajustado com ciclo anual e semi-anual.

#### Value:

A função retorna os valores de anomalias calculadas numa matriz "m.anomalia" cuja classe é uma matriz, salva num arquivo com extensão RData.

X: é uma matriz de dados com 3 dimensões (x,y,z)

T: é o período, o padrão é 12 meses.

rmNA: TRUE/FALSE. Se TRUE omitirá os valores NA's existentes na matriz de dado original. Se FALSE os NA's existente não serão omitidos.

file\_name: deve conter entre "aspas" o nome do arquivo a ser salvo com as anomalias calculadas

#### Warning:

rmNA = TRUE é padrão da função  
rmNA, T e file\_name são de preenchimento obrigatório conforme exemplo:  
mmq\_function(X,T=12,rmNA=TRUE,file\_name="anomalia\_result.RData")

#### Author(s):

Nancy Kazumi Taniguchi  
nancykazumi@usp.br

#### References:

Emery, W. J and Thomson, R. E.. 2001. Data analysis methods in physical oceanography. Elsevier Science.

#### Examples:

```
##criando matriz de dado SST 3D: lat, long, tempo em meses

#cria matriz 3d de dados nulo
m.sst <- array(data=NA, dim=c(10,15,300),dimnames=c("lat","long","mes"))

# preenchendo matriz 3d com dados
for (i in 1:10) #para cada posição de lat
```

```
{
  for (j in 1:15) #para cada posição de long
  {
    m.sst[i,j,] <- rnorm(300,22,3) #cria uma série de 300 valores de
distribuição normal com média 22 e desvio-padrão 3.
  }
}

#executa a função mmq_function
mmq_function(m.sst,T=12,rmNA=TRUE,file_name="anomalia_result.RData")
mmq_function(m.sst,T=12,rmNA=FALSE,file_name="anomalia_result.RData")
```

## Código da função MMQ

```
mmq_function <-
function(X,T=12,rmNA=TRUE,file_name="anomalia_result.RData")
{
  #cria matriz nula para guardar resultados de anomalia a ser calculado
  m.anomalia <- array(data=NA,
dim=c(10,15,300),dimnames=c("lat","long","mes"))
  resposta = "NA"
  ### tratando dados NA existente na matriz
  if (rmNA==TRUE)
  {
    m.sst <- (na.omit(X))
  }
  else
  {
    m.sst <- X
  }
  ###conhecendo a matriz m.sst
  nlat <- dim(m.sst)[1]
  nlong <- dim(m.sst)[2]
  nmes <- dim(m.sst)[3]
  ###definindo parâmetros dos ciclos
  T=12 #período 12 meses
  t=c(1:nmes) #nro de meses
  wt=(2*pi/T) #oscilação anual
  ones<-matrix(data=1,nrow=nmes, ncol=1, dimnames = NULL)
  X=cbind(ones, t, sin(wt*t),cos(wt*t),sin(wt*t*2),cos(wt*t*2)) #matriz
de coeficientes 1
  ##pagina 239 - Multivariate regression - Statistical Methods and Error
Handling - Emery & Thomson (2004)
  ## Data Analysis Methods in Physical Oceanography - Second and Revised
Edition
  for (i in 1:nlat) #para cada posição de lat
  {
```

```
for (j in 1:nlong) #para cada posição de long
{
  Y <- m.sst[i,j,] #selecionando serie temporal de uma posição
lat e long
  B=(t(X) %*% X)^(-1)%*%(t(X) %*% Y);
yh=B[1]+B[2]*t+B[3]*sin(wt*t)+B[4]*cos(wt*t)+B[5]*sin(2*wt*t)+B[6]*cos(2*wt*
t)
  anomalia=Y-yh
  m.anomalia[i,j,]=anomalia;
}
}
save(m.anomalia, m.sst, file = file_name)
resposta = paste("Matriz com resultados de anomalias foram salvos
no arquivo indicado",file_name)
return(resposta)
}
```

## Arquivos para download - função MMQ

Caso haja necessidade, você pode fazer o download da função MMQ (mmq\_function), o help , arquivo para teste da função ou solicitar esclarecimentos pelo email nancykazumi@usp.br

[funcao\\_mmq.r](#)

[help\\_similaridade.txt](#)

[teste\\_mmq\\_function.r](#)

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

Permanent link:  
[http://ecor.ib.usp.br/doku.php?id=05\\_curso\\_antigo:r2013:alunos:trabalho\\_final:nancykazumi:start](http://ecor.ib.usp.br/doku.php?id=05_curso_antigo:r2013:alunos:trabalho_final:nancykazumi:start)

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