

## Contextualização

A dinâmica populacional pode ser definida como a descrição na mudança no número de indivíduos de uma população ao longo do tempo e no espaço e dos fatores que influenciam essa flutuação populacional. Estes fatores podem ser classificados como: (1) bióticos e abióticos; (2) denso-dependentes e denso-independentes; (3) externos e internos dependendo o autor mas que em últimas afetam o número de indivíduos e o tempo de maturação, o que influencia o sucesso reprodutivo. Estudos populacionais são poucos devido a grande tempo que leva estudar como varia a densidade populacional de uma espécie. No entanto, estes estudos são de vital importância porque permite desde conhecer como eliminar uma “praga” de uma cultura, até conhecer o impacto da mudança climática nos diferentes ecossistemas.

## Ideia

A minha ideia é correlacionar dados quantitativos de várias variáveis com a abundância mensal de uma população para saber se existe sincronia entre os eventos e se essas variáveis estão de fato afetando ou não a flutuação populacional da espécie de estudo.

## Função

A **função** ira medir o **grau de relação linear entre duas variáveis quantitativas**. Em primeira instância, a função ira fazer um **teste de normalidade** para saber se os dados devem ser tratados com o **(1) coeficiente de correlação de Pearson** (Dados com distribuição normal) ou com o **(2) coeficiente de correlação de Spearman** (Dados não paramétricos). Posterior ao teste de normalidade, a função ira fazer o teste de Pearson ou Spearman segundo corresponda.

Estes dois testes variam entre os valores -1 e 1. O valor 0 (zero) significa que não há relação linear, o valor 1 indica uma relação linear perfeita e o valor -1 também indica uma relação linear perfeita mas inversa, ou seja quando uma das variáveis aumenta a outra diminui. Quanto mais próximo estiver de 1 ou -1, mais forte é a associação linear entre as duas variáveis.

## Objeto de entrada e argumentos

Os objetos e argumentos de entrada da função serão os seguintes: 1. Os dados serão um **data-frame com valores numéricos**, onde se espera que a primeira coluna seja a **abundancia mensal da espécie** de estudo e as outras colunas serão as diferentes **variáveis registradas (mensalmente)** que poderão ou não influenciar a **flutuação populacional** da espécie de estudo. 2. Os **argumentos** serão: (1) Entrada do **data-frame com os dados brutos** (2) **escolha do usuário** se vai analisas os dados com o coeficiente de Pearson ou Spearman. (3) **exclusão de valores faltantes (NA)**.

Posteriormente, para saber se existe sincronia ou não entre os eventos, será utilizada defasagem temporal com até dois meses.

## Objeto de saída

Finalmente, a função ira retornar um data-frame com as seguintes colunas: (1) coluna mostrando que **variável esta sendo correlacionada com a variável resposta** (abundancia da espécie); (2) o **número de meses de defasagem**; (3) coluna com os **valores numéricos indicando o valor de correlação** (-1 até 1).

Olá German,

A princípio sua função é muito simples, mas ela pode ser melhorada. A sua proposta foi basicamente fazer um teste de normalidade e uma correlação entre uma variável resposta e algumas variáveis explicativas quantitativas. Considerando que você está pensando em dados temporais (mensais), acho que pode ser interessante trabalhar também com estatísticas circulares. Vamos com calma, vou tentar explicar o que eu estou sugerindo.

Antes de tudo, acho que você podia pensar na sua função de forma mais geral. Por exemplo, uma função que verifica a sincronia entre uma variável resposta temporal e outras variáveis explicativas também temporais. Dessa forma, acho que apenas um teste de correlação é algo muito simples para uma função que pode ser muito mais legal. Os testes circulares são estatísticas que analisam justamente dados mensais e testam se existe alguma tendência temporal naqueles dados.

Veja esse artigo como um exemplo de como a estatística circular pode ser utilizada:

[http://www.scielo.br/scielo.php?script=sci\\_arttext&pid=S1519-69842009000400017](http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1519-69842009000400017)

Além disso, dê uma estudada nos pacotes “circular” e “plotrix”.

Na sua função, pode ser interessante gerar esses gráficos circulares tanto para a variável resposta como para as explicativas e verificar se houve sincronia temporal.

Outra questão que você precisa trabalhar melhor na sua proposta é a questão do tempo de defasagem. Essa ideia surgiu lá na entrada dos dados sem você ter dito nada sobre isso anteriormente. Eu entendo que a defasagem seria você verificar a correlação da variável resposta com as variáveis explicativas alguns meses antes, certo? Porém, você não me explicou nada disso e como isso entraria na sua função. Acho bastante interessante, então tente explicar melhor.

Algumas questões sobre o input:

- Acho que você podia colocar de forma mais geral o tipo de data-frame que a sua função irá aceitar. Por exemplo, um data-frame em que a primeira coluna seja a variável resposta temporal e as colunas seguintes sejam as variáveis explicativas temporais.

- Não entendi porque terá um argumento para escolher a correlação se a sua função irá determinar o tipo de correlação de acordo com os dados. Pense melhor nisso.

-Pensei se não seria mais interessante ter um argumento em que a pessoa pudesse escolher o tempo de defasagem. Assim, a sua função calcularia a correlação direta e a correlação com o tempo de defasagem solicitado. O que acha?

Algumas questões sobre o output:

- Achei interessante que a saída seja um data.frame. Porém, acho que faltam algumas coisas. Pense que cada linha do seu data.frame diz respeito a uma variável explicativa e você quer saber o valor de correlação direta e o valor de correlação com a defasagem temporal.

- Além disso, pense nas estatísticas e nos gráficos circulares.

Assim, acho que a sua função pode ficar mais interessante.

Precisamos fechar a sua proposta ainda. Aguardo a sua resposta. Por favor, me avise no e-mail abaixo assim que você me responder.

—[Leticia Zimback](#)

## \*\*Proposta A Second Round\*\*

Seguindo as dicas sobre a minha proposta, a função realizara o seguinte:

- Ver se há sincronia entre uma variável resposta temporal e outras variáveis explicativas temporais.
- Retornar um gráfico circular com os dados introducidos do data-frame.
- Retornar o valor de correlação linear (Pearson ou Spearman) além do valor da estatística circular (Rayleigh para dados com normalidade circular e Kuiper para não normalidade circular) para saber se existe uma tendencia dos dados hacia um mês no ano.

### **Objeto de entrada e argumentos**

Os objetos de entrada e argumentos da função serão os seguintes: 1. Os dados serão um **data-frame**, onde se espera que a primeira coluna seja de classe categorica contendo o ano ou os anos de estudo, a segunda coluna sera categorica contendo os meses, a terceira coluna sera a **variável resposta temporal** numerica, quarta quinta e sexta coluna serão **variáveis numéricas temporais** (por exemplo, temperatura, precipitação, etc.) que serão correlacionadas com a variável resposta. Todas as colunas deveram ter a mesma longitud de dados.

2. Os **argumentos** serão: (1) Entrada do **data-frame com os dados** brutos (2) **exclusão de valores faltantes** (NA).

### **Objeto de saída**

Finalmente, a função ira retornar um (1) data-frame com as seguintes colunas: (1) primeira coluna

mostrando que **variável preditoria esta sendo correlacionada com a variável resposta**; (2) segunda coluna mostrando o valor do coeficiente de correlação entre as variáveis (-1 até 1).

(2) lista com os valores dos testes circulares de Rayleigh ou Kuiper segundo corresponda para identificar tendencias dos dados ou homogeneidade.

(3) Gráfico circular dos dados da variável resposta indicada por meses

Olá German,

Melhorou muito!!! Do jeito que você organizou suas ideias acho que irá conseguir fazer um bom trabalho. Tenho algumas poucas questões somente para você pensar enquanto estiver fazendo a função:

Entrada da função

- Você colocou 4ª, 5ª e 6ª coluna como variáveis numéricas temporais explicativas, acho que você **NÃO** deve limitar o número de colunas na função. Assim, qualquer pessoa pode usar a função, seja com uma ou dez variáveis explicativas.

Saída da função

- Não entendi muito bem a primeira coluna do seu data-frame. Seria uma coluna com os nomes das variáveis explicativas que foram correlacionadas com a resposta?

- Acho que seria interessante fazer um gráfico circular para cada uma das variáveis explicativas também. Dessa forma você conseguiria verificar visualmente se há sincronia entre a variável resposta e cada uma das explicativas e não somente através da correlação.

Agora acho que você vai conseguir caminhar bem.

BOM TRABALHO.

—[Leticia Zimback](#)

## Proposta A Final Version

Seguindo as dicas e comentarios que a Leticia me falou pelo email sobre a minha proposta, a versão final da função final sera a seguinte:

**(A)** Ver se há sincronia entre uma variável resposta temporal e outras variáveis explicativas

temporais. Aqui a função retornara o valor de correlação linear (**Spearman**). Segundo a literatura, o coeficiente de correlação de Spearman é o mais robusto devido que ele não tem o pressuposto de linearidade entre as variáveis correlacionadas

[https://pt.wikipedia.org/wiki/Coeficiente\\_de\\_correla%C3%A7%C3%A3o\\_de\\_postos\\_de\\_Spearman](https://pt.wikipedia.org/wiki/Coeficiente_de_correla%C3%A7%C3%A3o_de_postos_de_Spearman).

**(B)** Retornar um gráfico circular com os dados introducidos do data-frame. O gráfico sera feito para cada **variável preditora** e para a **variável resposta**. Os gráficos terão a linha do vetor medio a qual indica qual é a direção dos dados dentro do gráfico circular e seu comprimento indicara que tão forte é esa direção.

**(C)** Valor da estatística circular para saber se existe uma tendencia ou direção dos dados hacia um mês no ano (sera feito para todas as variáveis):

(1)Primeiro passo é realizar um teste de normalidade circular para saber se os dados tem ou não uma distribuição normal circular (análoga com a distribuição normal de campana de Gauss). Para isso sera realizado um **teste de Kuiper** (análogo com o Teste de Kolmogorov-Smirnov) de distribuição normal circular (Houdson & Keatley 2009).[[[https://pt.wikipedia.org/wiki/Teste\\_de\\_Kuiper](https://pt.wikipedia.org/wiki/Teste_de_Kuiper)]]. Valores do teste acima do valor crítico, indicam a existência de um padrão nos dados.

(2) Posteriormente, a função realizara um **teste de Rayleigh** para saber se existe uma direção ou tendencia dos dados e se essa tendencia é significativa ou não. Valores inferiores no  $p = 0.05$  indicam que existe uma tendencia dos dados hacia um mês no ano.

### **Objeto de entrada e argumentos**

Os objetos de entrada e argumentos da função serão os seguintes:

1. Os dados serão um **data-frame**, onde se espera que a **primeira coluna** seja de classe categórica contendo o ano ou os anos de estudo, a **segunda coluna** sera categórica contendo os meses, a **terceira coluna** sera a variável resposta temporal numérica, **quarta quinta e sexta coluna** serão variáveis numéricas temporais (por exemplo, temperatura, precipitação, etc.) que serão correlacionadas com a variável resposta. Todas as colunas deveram ter a mesma longitud de dados.

2. Os argumentos serão: (1) Entrada do **data-frame com os dados brutos** (2) exclusão de **valores faltantes** (NA).

### **Objeto de saída**

Finalmente, a função ira retornar:

(1) **data-frame** com as seguintes colunas: **primeira coluna** mostrando que variável preditoria esta sendo correlacionada com a variável resposta; **segunda coluna** mostrando o valor do coeficiente de correlação entre as variáveis (valores desde -1 até 1). Exemplo

```
$`Coeficiente de correlação de Spearman da variavel resposta`
  variavel.correlacionada      valor.teste
```

1	temperatura	0.61844683
2	precipitacao	0.73047641
3	presas1	0.02448651

(2) **lista** com os seguintes valores:


1. Media dos valores numéricos de todas as variáveis (resposta e preditoras) por mês.
2. Valores dos testes circulares de Rayleigh, Kuiper e comprimento do vetor medio (que indica a força da direção dos dados).

(3) **Gráfico circular** dos dados da variável resposta e preditoras indicada por meses.

(4) Uma **mensagem de warning** falando como devem ser introducidos os dados para que a função rode corretamente.

Hudson, I. L., & Keatley, M. R. (2009). Phenological research: methods for environmental and climate change analysis. Springer Science & Business Media.

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

Permanent link:  
[http://ecor.ib.usp.br/doku.php?id=05\\_curso\\_antigo:r2016:alunos:trabalho\\_final:germanvillanueva9:proposta\\_a](http://ecor.ib.usp.br/doku.php?id=05_curso_antigo:r2016:alunos:trabalho_final:germanvillanueva9:proposta_a) 

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