

Mestranda pelo Programa de Pós-graduação em Ecologia e Evolução, UNIFESP.

O título da minha tese é “Efeito negativo da planta berçário *Guapira opposita* sobre espécies herbáceas em uma restinga arbustiva”, orientada pela Camila Castanho.

Meus exercícios

[exercicio_preparatorio.r](#)

[exercicio4.r](#)

[exercicio_graficos.r](#)

[exercicio_7a.r](#)

[exercicio_7b.r](#)

[exercicio_8.r](#)

[exercicio_9.2.r](#)

Proposta principal

A - Fumar ou se aposentar?

Contextualização:

O tabagismo atinge cerca de 10% da população brasileira, sendo mais frequente entre os jovens universitários, chegando a 21,6%. Esse hábito é nocivo à saúde e pode desencadear diversas doenças, como por exemplo: doenças respiratórias e cardiovasculares, câncer, úlcera, osteoporose, catarata, impotência sexual no homem e infertilidade na mulher. Uma forma de acabar esse mau hábito, é parar de fumar regressivamente. Ou seja, diminuir x cigarros ao mês, de forma que seu corpo se acostume com a quantidade reduzida de nicotina. Decidiu parar de fumar? Bom, agora que você finalmente optou por uma vida mais longa e saudável é preciso começar a se preocupar com seu futuro e sua aposentadoria. A partir de hoje, para cada centavo que antes era gasto em cigarro, você irá aplicá-lo no tesouro selic. Diante das incertezas que vivemos atualmente, será que agora rola a tão sonhada aposentadoria?

Planejamento:

Pré-função:

Criar um dataframe contendo os valores de 20 cigarros (1 maço) de acordo com:

1. Marca: marca $\leftarrow$ c (“Camel”, “Chesterfield”, “Dallas”, “Derby”, “Downtown”, “Dunhill”, “Eight”, “Free”, “Hilton”, “Hollywood”, “Kent”, “L&M”, “Lucky strike”, “Marlboro”, “Minister”, “Shelton”, “Winston”)
2. Versão: versao $\leftarrow$ c (“azul”, “prata”, “vermelho”)
3. Embalagem: embalagem $\leftarrow$ c (“maleavel”, “rigida”)
4. Estado do Brasil estado $\leftarrow$ c (22 estados do brasil)

Entrada:

aposentadoria ← function (data, apos.idd = 60, infl = 0.0603, selic = 0.065)

data = vetor contendo nesta ordem:

1. ncig (número de cigarros fumados diariamente)
2. stopcig (número de cigarros diminuídos por mês)
3. marca
4. versao
5. embalagem
6. estado
7. sexo
8. idade - idade do fumante (em anos com 1 cada decimal. EX: 20 anos e 6 meses = 20,5)

apos.idd = idade de aposentadoria. Por default, a aposentadoria é de 60 anos (mulher). if M, apos.idd = 65

infl = inflação sobre o preço do cigarro. Em 2018 esse valor foi de 0.0603. Portanto, por default, a inflação ao ano é de 0.0603

selic = por default a taxa selic é igual a taxa referente ao ano de 2018

Verificações:

1. if marca != das marcas listadas: mean do preço de todas as marcas
* Warning avisando o que será calculado
2. if versao != das versoes listadas: mean das opções dentro da marca apresentada
* Warning avisando o que será calculado
3. if embalagem != das listadas: mean das opções dentro da marca e versao apresentada
* Warning avisando o que será calculado
4. if sexo == "M": apos.idd=65
5. if class(idade) != numeric
* Stop (call = "A idade precisa ser da classe 'numeric'!)
6. if estado != dos listados: mean do preço do cigarro em todos os estados do brasil
* Warning avisando o que será calculado

Pseudo-código:

1. calcular o numero de cigarros consumidos por mês

$\text{ncig.mes} \leftarrow \text{ncig} * 30$

2. preço do cigarro unitario

$\text{cig.unid} \leftarrow \text{preco} / 20$

3. multiplicar o preco unitario pela quantidade consumida no mes

$\text{preco.mes} \leftarrow \text{ncig.mes} * \text{cig.unid}$

4. calcular o tempo (em dias) para parar totalmente de fumar

$\text{abstinencia} \leftarrow \text{ncig.mes} - \text{stopcig} * 30$

5. calcular o valor investido (inv) desde o t0 até o t abstinencia

a. por exemplo: no t0 investe o valor vezes stopcig

b. no t abstinencia investe o valor de ncig

c. a partir do t abstinencia passa a investir o mesmo valor mensalmente até aposentar

6. associado ao valor investido é calculado o rendimento do tesouro selic ao mês

7. valor investido + rendimento:

$\text{inv} + \text{inv} * (\text{selic} / 12)$

8. apos 12 meses o valor do cigarro sofre um reajuste de infl

a. $\text{preco atualizado} = \text{cig.unid} + \text{cig.unid} * \text{infl}$

9. isso se repete x vezes (for 1 : x)

a. diferença entre idade atual e a idade de aposentadoria ($\text{diff}() = x$)

10. plot com o valor investido ao longo do tempo

a. eixo x = tempo

b. eixo y = valor investido

Arquivo de saída:

1. valor total em reais do dinheiro investido com seu devido rendimento

2. plot com o valor investido ao longo do tempo

¹<https://www.inca.gov.br/observatorio-da-politica-nacional-de-controle-do-tabaco/dados-e-numeros-pr-avalencia-tabagismo>

Comentários Pamela Santana

Oi Renata!

Adorei sua proposta A! A ideia tá muito interessante e bem elaborada. Tem algumas coisinhas que podemos pensar:

i) Acho que vale pensar bem em como será sua entrada de dados. Por exemplo, se você entra com todos os dados em um vetor, como você fará cada uma das suas verificações? Você dirá ao R exatamente a posição no vetor que deve ser testada?

ii) *stopcig* já é a quantidade de cigarros diminuída no mês né? Porque você multiplica por 30 novamente para calcular abstinência?

iii) Que tal se você permitisse que o usuário da função escolhesse outras opções de investimento?

B- Lotofácil:

Contextualização:

O sonho de ganhar na loteria afeta todos nós. Há quem diga que é impossível e há quem diga que existe uma fórmula para atingir esse objetivo. Trataremos aqui, do jogo da Lotofácil, que consiste em escolher 15 números entre os 25 disponíveis (de 1 a 25). Abaixo podemos ver uma tabela contendo as premiações dado o número de acertos.

Número de acertos e premiação

11 pontos - R\$ 4.00

12 pontos - R\$ 8.00

13 pontos - R\$ 20.00

14 pontos - 20% da arrecadação do concurso

15 pontos - 65% da arrecadação do concurso

É importante lembrar que nunca foram sorteados dois ou mais jogos idênticos. Ou seja, cada combinação de 15 números só foi sorteada uma vez em toda a história da lotofácil. Se você quiser investir mais dinheiro a fim de aumentar a probabilidade de ganhar, saiba que é possível escolher mais de 15 números, dentre os 25 disponíveis.

Quantidade de números e valor

15 - R\$ 2.00

16 - R\$ 32.00

17 - R\$ 272.00

18 - R\$ 1632.00

Planejamento:**Arquivos de entrada:**

lotofacil ← function (anteriores, data, n=15, vezes=10)

1. Anteriores = arquivo contendo todos os concursos que já aconteceram (ainda estou procurando esse arquivo)
2. Data = vetor contendo os números que saíram no sorteio mais recente
3. n = quantidade de números escolhidos pelo usuário para realizar o jogo. Por default, n=15
4. vezes = número de jogos que o usuário deseja fazer. Por default o número de jogos é igual a 10.

Verificações:

1. If data != as.vector

Stop (call = "o objeto de entrada precisa ser um vetor")

2. If data != numeric

Stop (call = "o vetor precisa ser da classe 'numeric'")

3. 'n' precisa ser um valor entre 15 e 18

4. 'anteriores' não pode ter NA

5. 'anteriores' só pode conter números de 1 a 25

Pseudo-código

Inicialmente serão levados em conta os seguintes critérios:

1. Quantas vezes cada número já foi sorteado
2. Relação de números pares e ímpares
3. Quantidade de números primos
4. Quantidade de números que são múltiplos de 3
5. Quantidade de números de Fibonacci
6. Soma das dezenas
7. Número de dezenas repetidas do concurso anterior

Para cada critério:

Calcular a proporção de ocorrência para cada evento do espaço amostral

Existe diferença significativa entre as proporções?

Se sim, continua a função levando em consideração este critério

Criar um objeto indicando quais são esses números (por exemplo: $\text{par} \leftarrow c(1,3,5,7,\dots)$)

Se não, a função não levará em conta este critério

Para o critério soma, caso a diferença seja significativa:

Calcular todas as combinações possíveis que resultem o intervalo de soma mais provável de ocorrer

Função for (i in 1:vezes)

Considerando todas probabilidades que foram significativamente maiores do que as demais

Além disso, esses números precisam ter uma similaridade com o concurso anterior de $x\%$

E a combinação gerada precisa ser diferente de todas as combinações que já ocorreram

Arquivos de saída:

1. Valores simulados para o número de jogos que o usuário deseja fazer (argumento = vezes) de acordo com o tipo de jogo que o usuário escolher ($n=15$ a $n=18$)
2. Histograma simulando os valores mais prováveis de sair.

Comentários Pamela Santana

Sua proposta B também está bem interessante! Mas, lembre-se que sua função deve realizar uma tarefa *geral* ou aplicável recorrentemente. Então, tenho alguns comentários:

i) Se eu entendi bem, da forma como ela está, ela é aplicável somente à Lotofácil e neste caso, o número de anteriores seria sempre o mesmo, né? Se for assim, uma vez calculada o padrão do passado, a função não precisa dessa entrada sempre. Se você resolver manter somente para Lotofácil, anteriores não precisa ser um argumento da sua função. No entanto, se você considerar que ao longo de anos os padrões podem ir mudando. Então, é válido. Mas... e se você tentar generalizar essa função para mais de um tipo de jogo de sorte. Que tal? Desde que o jogo seja com entrada de números.

ii) "Anteriores" são os sorteados para múltiplas quantidades de números (n)? Como esse dado está organizado? É um data frame? Ou são data-frames distintos a depender de n ?

ii) Ainda, se eu entendi bem, você quer fazer comparativos de probabilidades certo? Então, para verificar isso, você precisará ver no "data" todos os critérios que você olhou para os anteriores também. Por exemplo: qual a relação de números pares e ímpares média no passado? (isso você faz quando diz *Calcular a proporção de ocorrência para cada evento do espaço amostral*, mas como você vai calcular isso?) E a comparação com os valores de *data*? Você precisaria ter calculado para *data* também todos os critérios para poder comparar, certo? Qual era a chance de isso ocorrer dado um certo valor de p ?

iii) Não fica muito claro como você vai calcular usando a **soma** para verificar o intervalo mais provável de ocorrer. Você poderia explicar isso um pouquinho melhor, por favor?

iv) O que exatamente seu histograma vai mostrar? A frequência de uma dada combinação?

Eu sugeriria que você prosseguisse com sua **proposta A**, apesar de ambas estarem interessantes. A sua proposta B ainda tem que elaborar mais nos cálculos e nos comandos. Seu pseudo-código ainda dá uns saltos. Então, acredito que a implementação da A seria mais fácil.

Pense nesses pontos que levantei e me escreva se precisar de qualquer coisa: [Pamela Santana](#)

Um abraço!

[aposentadoria_funcao_final.r documentacao_aposentadoria.docx](#)

From:

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

Permanent link:

http://ecor.ib.usp.br/doku.php?id=05_curso_antigo:r2019:alunos:trabalho_final:renatamarquesribeiro:start



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