Primeira lista de exercícios

Estatística Computacional

Exercício 1. Crie os seguintes vetores:

(a) (10, 11, 12,..., 30)

(b) (30, 29, 28,..., 10)

(c) (10, 11, 12,..., 30, 29, 28,..., 10)

Exercício 2. Use a função `help` do R para descobrir o funcionamento das funções `rep` e `seq`. Em seguida, utilize estas funções para resolver os seguintes itens:

(a) Crie o vetor (2, 4, 6, 8, 2, 4, 6, 8,..., 2, 4, 6, 8), em que há dez ocorrências do número 2.

(b) Crie o vetor (2, 4, 6, 8, 2, 4, 6, 8,..., 2, 4, 6, 8, 2), em que há onze ocorrências do número 2 e dez ocorrências dos números 4, 6 e 8.

Exercício 3. Utilize a estrutura de vetores do R para realizar as seguintes somas:

(a) $\sum\limits_{n=20}^{30}(n^2+4n)$;

(b) $\sum\limits_{n=10}^{20}(\frac{3^n}{n}+\frac{2^n}{n^2})$.

Exercício 4. Numa urna há bolas idênticas numeradas de 1 até 100. Serão extraídas 40 bolas com reposição desta urna. Simule este experimento e guarde o resultado dos sorteios em um vetor.

(a) Quantas bolas pares foram sorteadas?

(b) Quantas bolas maiores do que 70 foram sorteadas?

(c) Em quais retiradas (posições) foram sorteadas as bolas ímpares?

Observação: em R, o operador `%%` retorna o resto da divisão euclidiana entre dois inteiros; dessa forma, por exemplo:

```
9 %% 2
```

```
[1] 1
```

```
20 %% 8
```

```
[1] 4
```

Exercício 5. Crie um função no R que irá simular sucessivos lançamentos de um dado até que o número 4 seja obtido pela segunda vez. A função deverá retornar o número de lançamentos que foram necessários até o 4 ser obtido pela segunda vez. Assim, se os sorteios foram 3, 6, 6, 5, 4, 2, 4 a função deverá retornar 7.

Exercício 6. Utilize a função do exercício anterior para replicar o experimento dez mil vezes. Para cada replicação, guarde o número de lançamentos num vetor chamado `quantidades`. Por fim, calcule a média de `quantidades`. Interprete o resultado obtido.

Exercício 7. Os dois primeiros termos da sequência de Fibonacci são iguais a 1. Os termos subsequentes da sequência são encontrados somando os dois termos imediatamente anteriores. Escreva uma função com parâmetro de entrada `n` chamada *fibonacci* que retornará os primeiros `n` termos da sequência de Fibonacci para qualquer $n \geq 3$. Exemplo:

```r
fibonacci(10)
```

```
 [1]  1  1  2  3  5  8 13 21 34 55
```

Exercício 8. Michael Scott é gerente regional da empresa Dunder Mufflin. Para as festividades de fim de ano, Michael propôs aos funcionários Dwight Schrute, Jim Halpert, Kevin Malone e Creed Bratton a realização de um *amigo oculto* entre eles. Consideraremos que o sorteio do amigo oculto deu errado quando uma pessoa sortear ela mesma (Michael tira Michael, por exemplo). Simule o sorteio do amigo oculto. Se ele deu certo, atribua o valor 1; caso contrário, atribua o valor 0 (zero). Em seguida, replique este experimento cem mil vezes e calcule a proporção de vezes que o amigo oculto deu errado.

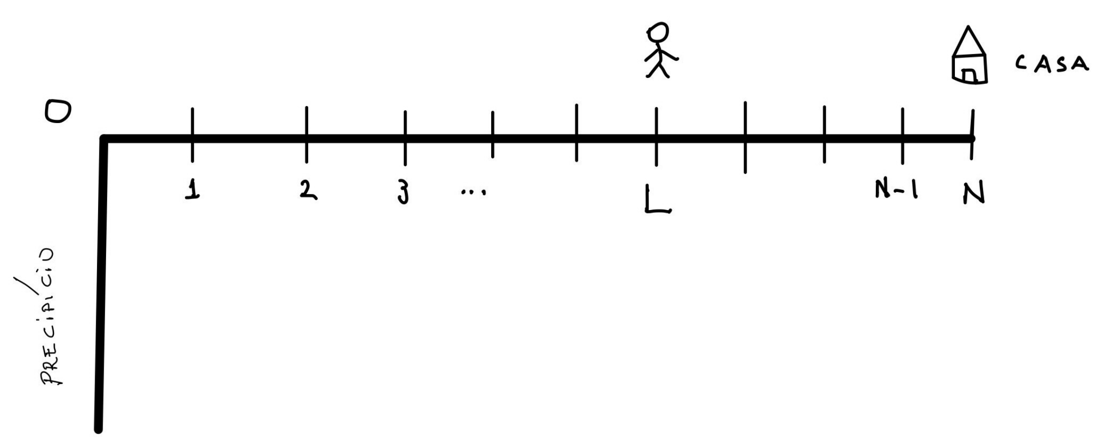
Exercício 9. No jogo de *Craps* dois dados são lançados:

- se a soma for 7 ou 11, você ganha o jogo;
- se a soma for 2,3 ou 12, você perde o jogo;
- caso contrário, os dois dados são rolados novamente até obter-se 7 (você perde) ou até obter-se a soma inicial (você ganha).

Simule uma partida do jogo de Craps. Em seguida, replique este experimento 100 mil vezes; para cada experimento, atribua 1 para uma vitória e zero para uma derrota. Calcule a proporção de vezes que você ganhou o jogo.

Exemplo: as seguintes sequências (cada entrada é a soma dos dois dados) resultam em vitória: (7), (11), (5, 4, 5), (4, 5, 6, 12, 4); as seguintes sequências resultam em derrota: (2), (4, 11, 7), (8, 5, 2, 3, 9, 7).

Exercício 10. Luke Skywalker realizará o seguinte passeio aleatório na reta: a reta do passeio é formada pelos números inteiros de zero até N; Luke está em um ponto L que é maior do que zero e menor do que N; Luke lança uma moeda honesta; se sair coroa, ele dá um passo para a esquerda (e termina na posição $L-1$ da reta); se sair cara, ele dá um passo para a direita (e termina na posição $L+1$ da reta). Luke continuará a lançar a moeda e se deslocará até que ele chegue em sua casa (e lá ele vai dormir e o passeio acaba) ou até que ele chegue (caia) no precipício (e, óbvio, o passeio também acaba nesse caso).

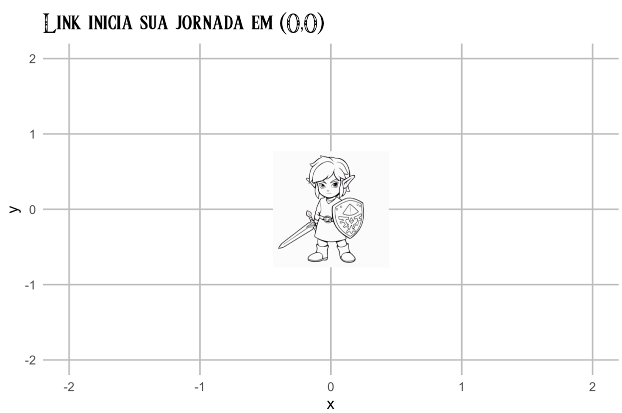


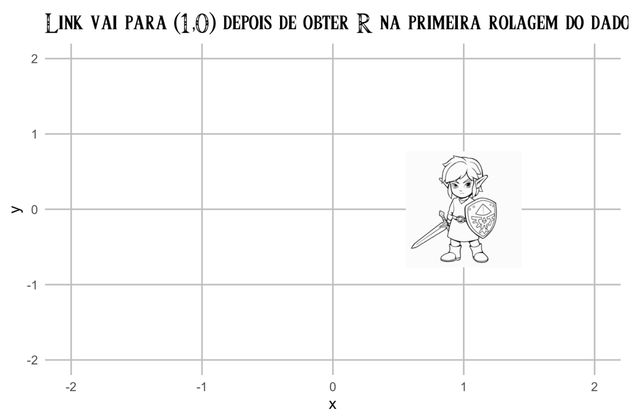
(a) Para $N = 20$, crie uma função cuja entrada seja L (um número maior do que zero e menor do que 20) e que retorne 1 se Luke terminou um passeio em sua casa ou retorne zero se Luke caiu no precipício.

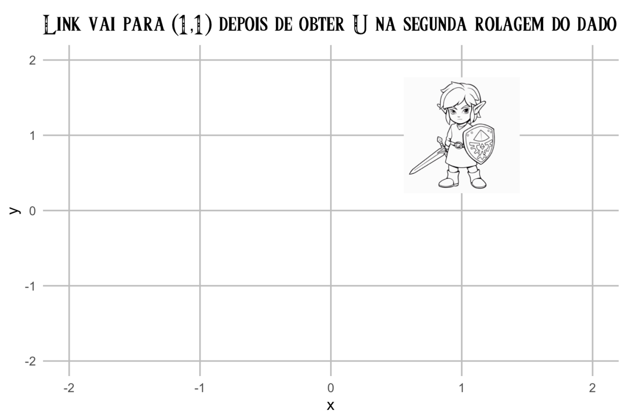
(b) Crie uma função cuja entrada seja L; esta função deverá replicar o passeio da letra (a) 10 mil vezes e retornar a proporção de vezes que Luke chegou em sua casa. Sugestão: crie um vetor que, para cada replicação, guardará o resultado de um passeio; cada entrada deste vetor será zero ou 1; zero se Luke caiu no precipício e 1 se Luke chegou em casa.

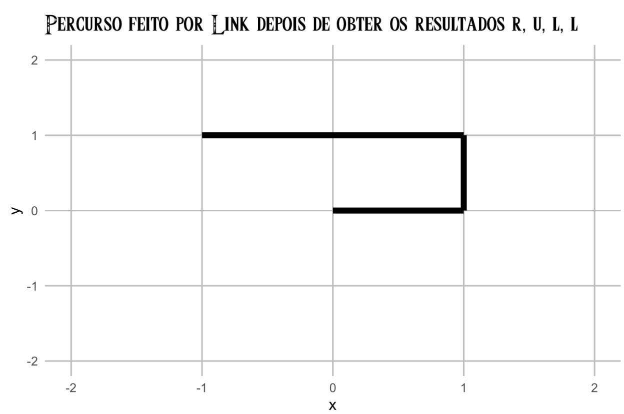
(c) Use a função criada em (b) para $L = 1, 2, \ldots, 19$ e, em seguida, use esses valores para plotar um gráfico de $x = 1 : 19$ por y, em que y são as proporções retornadas pela função criada em (b) para cada x.

Exercício 11. Link desperta após um sono de cem anos e encontra o reino de Hyrule em ruínas, necessitando recuperar suas memórias e enfrentar o mal desencadeado por Calamity Ganon. Desprovido de suas lembranças, Link adota um método peculiar para explorar o reino: lançando um dado de 4 lados. As faces deste dado são rotuladas como L, R, U e D. A direção de cada passo que Link dá é determinada pelo resultado do dado que ele lança: se sair L ele dá um passo para a esquerda; se sair R, ele dá um passo para a direita; se sair U, ele dá um passo para cima; se sair D, ele dá um passo para baixo. Suponha que Link comece sua jornada no ponto $(0, 0)$ e que os quatro primeiros lançamentos do dado foram R, U, L e L. Portanto, a sequência de passos de Link, neste caso, é: $(0, 0) \to (1, 0) \to (1, 1) \to (0, 1) \to (-1, 1)$. As Figuras 1, 2 e 3 apresentam, respectivamente, Link no início de sua jornada, após o primeiro lançamento do dado e após o segundo lançamento do dado; a Figura 4 apresenta o percurso total feito por Link após os quatro lançamentos.









(a) Simule um passeio de 8 passos com Link começando na origem. A saída desta simulação deve apresentar o ponto do plano em que Link está após os 8 passos.

(b) Agora replique dez mil vezes o experimento de (a) e determine a proporção de vezes em que Link retornou para a origem depois de 8 passos. Comente o que esta proporção significa.

(c) Escreva uma função em que o usuário entra com um número N: se o número N for ímpar, então a função retorna: "impossível retornar a origem depois de um número ímpar de passos" (quando estiver estudando esta questão, se convença que 'isso é verdadeé impossível regressar à origem em um número de passos ímpar); se o número N for par, então a função retorna uma frase informando a proporção de vezes em que Link retornou a origem depois de N passos ao se realizar o experimento dez mil vezes.

Exercício 12. Considere o seguinte jogo: Steven e Garnit escolherão, cada um, uma sequência de tamanho 3 em que cada entrada da sequência é cara ou coroa; logo em seguida, uma moeda será lançada três vezes; se aparecer a sequência de um dos jogadores, este jogador vence e o jogo acaba; caso não apareça a sequência de nenhum deles, a moeda é lançada pela quarta vez e os três últimos lançamentos são analisados; se nestes três últimos lançamentos aparecer a sequência de um dos jogadores, este jogador vence e o jogo acaba. Se isto não acontecer, a moeda é lançada pela quinta vez e os três últimos resultados são analisados; se aparecer a sequência de um dos jogadores, este jogador vence e o jogo acaba. Este processo é realizado até que apareça a sequência que um dos dois escolheu; se aparecer primeiro a sequência de Steven, ele ganha; se aparecer primeiro a sequência de Garnit, ela vence. Convencione que cara seja 1 e que coroa seja zero. Supondo que Steven escolheu a sequência (0, 1, 0) e que Garnit escolheu a sequência (0, 0, 1), simule uma partida deste jogo. A simulação deve retornar `steven` caso Steven tenha vencido ou deve retornar `garnit` caso contrário. Replique o experimento 10 mil vezes e calcule a média de vitórias de Garnit. Comente o resultado obtido. (6 pontos)

Observação: Suponha que os três primeiros lançamentos foram (1,0,0). Logo, ninguém ganhou e a moeda é lançada pela quarta vez. Suponha que o quarto lançamento foi 0; logo os três últimos lançamentos foram (0,0,0) e ninguém ganhou. Na quinta vez saiu 1 e, portanto, os três últimos lançamentos foram (0,0,1) e o jogo acaba com vitória de Garnit. As sequências (0, 1, 0), (1, 0, 1, 0) e (1, 1, 0, 1, 0) fazem Steven vitorioso; as sequências (0, 0, 1), (0, 0, 0, 1) e (1, 0, 0, 0, 1) fazem Garnit vitoriosa.

Exercício 13. Harold Frederick Shipman (Nottingham, 14 de janeiro de 1946 — Wakefield, 13 de janeiro de 2004), conhecido como “Doutor Morte”, foi um médico e assassino em série britânico condenado pela morte de muitos pacientes entre as décadas de 1970 e 1990. Dr. Shipman é, talvez, o assassino em série mais prolífico da História Moderna. O arquivo `dados.txt` contém informações sobre o sexo, a idade, o local da morte (casa do paciente; hospital; casa de repouso) e o ano da morte das vítimas de Shipman. Antes de responder as questões abaixo, abra o arquivo `dados.txt` e compreenda sua estrutura. Importe o arquivo para o R e utilize-o para responder os seguintes itens.

(a) Escolha um gráfico apropriado para representar as frequências das categorias da variável sexo. Comente os resultados encontrados.

(b) Apresente o histograma da variável idade em 8 (argumento `bins` na geometria do histograma) intervalos. Comente os resultados obtidos. Analise este gráfico para cada gênero.

(c) Apresente o boxplot da variável idade. Comente os resultados obtidos.

(d) Apresente um gráfico para representar o local da morte. Comente os resultados obtidos.

(e) Analise graficamente o ano da morte das vítimas de Harold Shipman.

(f) Com base nas informações obtidas nos itens anteriores, escreva um parágrafo sobre o padrão e o perfil das vítimas de Harold Shipman.

Exercício 14. Os bonobos, também conhecidos como chimpanzés-pigmeus, são encontrados apenas em uma área limitada da República Democrática do Congo, na África Central. Eles são conhecidos por sua sociedade matriarcal, onde as fêmeas têm um papel dominante na tomada de decisões e na resolução de conflitos. Os bonobos são geralmente mais pacíficos do que os chimpanzés, e os conflitos dentro do grupo são frequentemente resolvidos através de interações sociais, como a partilha de alimentos e o sexo. Eles são altamente sociais e valorizam as relações e o contato físico com outros membros do grupo. Por outro lado, os chimpanzés têm uma distribuição mais ampla, encontrados em várias regiões da África Central e Ocidental. Eles têm uma sociedade mais hierárquica e dominada por machos, onde o poder é frequentemente obtido por meio de competição e agressão. Os chimpanzés também exibem comportamentos violentos e podem entrar em confrontos territoriais com outros grupos. Eles são conhecidos por usar ferramentas, como galhos para pescar cupins e pedras para quebrar nozes.

O conjunto `primatas.txt` apresenta informações sobre tamanho (centímetros), peso (libras) e gênero de bonobos e de chimpanzés. Abra o arquivo e veja como ele está organizado.

(a) Importe o arquivo para o ambiente do R. Conheça sua estrutura e peça um resumo dos dados com alguma função. (1 ponto)

(b) Construa um gráfico de barras contando quantas espécies de bonobos e chimpanzés há no conjunto. Construa também um gráfico de barras mostrando a frequência de machos e fêmeas de cada espécie. (4 pontos)

(c) Construa um gráfico para comparar as fêmeas e os machos dos bonobos. Em seguida, construa, também, um gráfico para comparar as fêmeas e os machos dos chimpanzés. (4 pontos)

(d) Construa um gráfico para comparar as fêmeas dos bonobos e dos chimpanzés. Em seguida, construa também um gráfico para comparar os machos dos bonobos e dos chimpanzés. (4 pontos)

(e) A partir das análises dos itens anteriores, escreva um pequeno texto contendo informações sobre os bonobos e os chimpanzés, como exemplo: diferenças entre os gêneros de cada espécie e diferenças entre as espécies. (5 pontos)

(f) A partir das variáveis `tamanho`, `peso` e `genero`, construa um modelo de árvore de decisão utilizando estruturas condicionais que seja capaz de prever a espécie de uma observação. Calcule a acurácia do modelo. (10 pontos)