

PRIMEIRA PROVA DE ESTATÍSTICA - 01/11/2022

Questão 1. A roleta americana é composta por 36 casas: 18 pretas, 18 vermelhas e 2 verdes. Você aposta numa casa preta ou vermelha. Suponha que você comece o jogo com 10 mil dólares. Se você ganha, você recebe 100 dólares; se você perde, você paga 100 dólares. Assim, se as primeiras partidas resultaram em vitória, derrota, derrota, derrota e vitória, então as flutuações de seu dinheiro serão: 10000 (início), 10100, 10000, 9900, 9800, 9900.

- (a) Simule 200 rodadas do jogo; ao final das 200 partidas, exiba a quantia final do seu dinheiro. Considere que você sempre aposta nas casas vermelhas.
Sugestão: sorteie um número do conjunto 1:38; se sair um número menor ou igual a 18, você vence; caso contrário, você perde. (3 pontos)
- (b) Replique a letra (a) mil vezes; ao final da simulação você deve exibir um vetor de tamanho mil em que cada entrada desse vetor representa a quantia final em dinheiro para 200 rodadas consecutivas. Por fim, calcule a média desse vetor, isto é, a média do seu dinheiro final após 200 rodadas. (7 pontos)

Questão 2. Papagaio-do-mar é o nome comum dado às aves charadriiformes da família dos alcídeos, pertencentes ao gênero *Fratercula*. Existem três espécies de papagaios-do-mar conhecidas: *arctica*, *corniculata* e *cirrhata*. O conjunto **papagaio.csv** apresenta informações sobre o peso (em gramas), a envergadura da asa (em centímetros) e a espécie de 412 papagaios-do-mar.

- (a) Leia o arquivo **papagaio.csv**. Em seguida, comece a analisar os dados a partir das funções **str** e **summary**. (0.5 ponto)
- (b) Conforme pôde ser observado na parte (a), os dados estão ordenados a partir da variável **especie**. Embaralhe o conjunto de dados de modo a deixar a ordenação das observações aleatória. (1 ponto)
- (c) Se a variável **especie** for do tipo **character**, transforme-a em uma variável do tipo **factor**. (0.5 ponto)
- (d) Construa um gráfico de barras para a variável **especie**. Comente os resultados obtidos. (2 pontos)
- (e) Construa um histograma para a variável **peso** da espécie *arctica*. Comente os resultados obtidos. (2 pontos)
- (f) Determine a média, o desvio padrão e o coeficiente de variação para a variável **envergadura** para cada uma das espécies. Comente os resultados obtidos. (2 pontos)
- (g) Numa mesma janela, exiba os boxplots para o peso de cada uma das espécies. Comente os resultados obtidos. (2 pontos)
- (h) Construa um gráfico de pontos em que a variável **peso** está no eixo *x* e a variável **envergadura** está no eixo *y*. Cada espécie deve ser representada por uma cor diferente. (2 pontos)
- (i) Divida o conjunto de dados em dois: um para treino e outro para teste. O conjunto de treinamento deve conter 80% dos dados do conjunto inicial. (1 ponto)
- (j) Utilize a estrutura encontrada em (h) para criar uma árvore de decisão para classificar a espécie de um papagaio-do-mar a partir do peso e da envergadura de sua asa. (4 pontos)
- (k) Calcule a sua taxa de acerto para o modelo construído em (f) utilizando nesse cálculo o conjunto de teste. (1 ponto)