

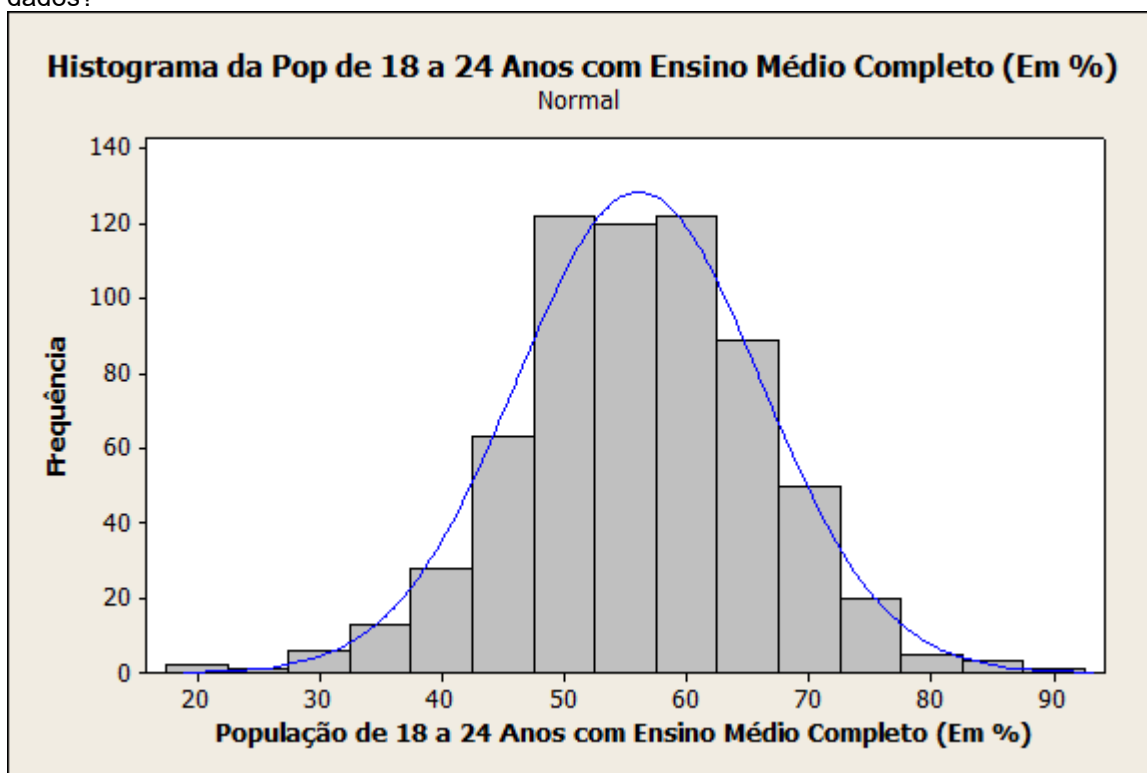
LISTA DE EXERCÍCIOS 3

Os exercícios 1 e 2 referem-se a base de dados “vendas mensais de carros”.

- 1) Qual a média e o desvio padrão da venda de carros durante o período?
- 2) Qual os valores obtido quando aplicadas as normalizações de dados linear e por desvio padrão?
- 3) Para esta base de dados, você utilizaria normalização linear ou por desvio padrão? Justifique sua resposta.

Os exercícios abaixo referem-se a base de dados “População de 18 a 24 anos com ensino médio completo - municípios SP”

- 4) O arquivo apresenta dados da População de 18 a 24 Anos com Ensino Médio Completo (Em %), em 2010, para 645 municípios. O histograma de frequências abaixo mostra que é razoável considerar que os dados têm distribuição normal. Considere que uma observação é um outlier se estiver a mais de 3 desvios-padrão distante da média. Quais são os municípios outliers deste conjunto de dados?



- 5) Explique como uma análise de discriminante pode construir fronteiras de decisão lineares.
- 6) Porque as funções discriminantes são classificadores paramétricos?
- 7) Dado as fronteiras de decisão abaixo, qual possui menor sensibilidade à outlier ou presença de ruído. Justifique sua escolha.
- 8) Porque o KNN é um classificador não-paramétrico?
- 9) O classificador KNN contrói fronteiras de decisão lineares ou não lineares? Justifique sua resposta.
- 10) Discuta como a escolha do valor de K pode influenciar o algoritmo.
- 11) Implemente o algoritmo KNN, LDA e QDA e apresente as taxas de acerto para as seguintes bases de dados.
 - a. Classificação da origem de amostras de vinhos: <https://archive.ics.uci.edu/ml/datasets/Wine>
 - b. Classificação do nível de consumo alcóico de estudantes a partir de informações sociais.
<https://archive.ics.uci.edu/ml/datasets/STUDENT+ALCOHOL+CONSUMPTION>