LISTA DE EXERCICIOS PARA ENTREGAR INDIVIDUALMENTE

LEMBREM-SE:

* IMPORTANTE EXPLANAR E EXPLICAR O ENTENDIMENTO NA SUA ENTREGA.
* O NOME DO ARQUIVO DEVE SER SEU **PRIMEIRONOME\_SOBRENOME** (Exemplo: Adriana\_Silva)
* VOCÊ FARÁ A ENTREGA DO TRABALHO EM WORD COM TODAS AS EXPLICAÇÕES (cole o resultado do software e interprete).
* NÃO USAR COR VERMELHA, POIS ELA FICA DESTINADA ÀS CORREÇÕES!!

1. (0,5) Explique o propósito de usar variáveis dummy em uma regressão linear com variáveis categóricas.

Regressões lineares exigem variáveis numéricas, tornando necessário transformar variáveis categóricas em um formato numérico adequado. Para isso, utilizamos variáveis dummy, que são representações binárias das categorias indicando a presença ou ausência da informação.

1. (0,5) Explique como a transformação logarítmica pode ajudar em um modelo de regressão linear.

Linearizar relações não lineares. Aplicar log() reduz a curvatura da relação.

Reduzir a influência de outliers, pois o log diminuiu mais os valores maiores mais do que valores menores e deste modo a influência de valores muito elevados diminui.

Também pode tornar a variabilidade mais homogênea.

1. (1,0) Explique o objetivo principal da transformação Box-Cox em um modelo de regressão linear.

Procura encontrar uma transformação ótima da variável dependente para atender pressupostos da regressão linear (Normalidade dos resíduos, variância constante e Linearidade entre variáveis preditoras e resposta) e tentar capturar melhor o comportamento de distribuição não lineares.

1. (1,0) Descreva o processo de escolha do parâmetro λ (lambda) na transformação Box-Cox.

Tentativa/experimentação. Diferentes valores de λ são testados de modo a encontrar aquele que maximiza a função de verossimilhança.

1. (1,0) Como a transformação Box-Cox ajuda a lidar com a heterocedasticidade em um modelo de regressão linear?

A transformação Box-Cox torna a variância mais homogênea, ajustando os dados para que os erros do modelo fiquem mais equilibrados. A variância dos resíduos da regressão não sendo constante ao longo dos valores preditos viola um dos pressupostos dos mínimos quadrados ordinários (OLS).

1. (1,0) Quais são as limitações da transformação Box-Cox e em que situações ela pode não ser apropriada?

Exige dados positivos: A função não é definida para valores y≤0 (não pode ser usada quando existem valores negativos ou zeros).

1. (1,0) Dê um exemplo de como aplicar a transformação Box-Cox em um conjunto de dados e interpretar os resultados.

Se estivermos com um conjunto de dados com a variável Valor\_Venda (alvo) e queremos ajustar um modelo de regressão linear, pois notamos que os resíduos não seguem uma distribuição normal e há heterocedasticidade.

Aplicamos a transformação na variável alvo (Valor\_Venda).

Para analisar os resultados temos que fazer a transformação inversa para chegarmos novamente no valor da variável Valor\_Venda.

1. (1,0) Descreva o método dos mínimos quadrados e como ele é utilizado na estimativa dos coeficientes de regressão.

O método dos mínimos quadrados ordinários (OLS) é a técnica usada para estimar os coeficientes da regressão linear. Ele minimiza a soma dos quadrados dos erros.

Ele busca minimizar a soma dos erros ao quadrado entre os valores reais (y) e os valores previstos pelo modelo (y\_chapeu), ou seja, ajustamos uma linha aos dados de forma que os erros sejam os menores possíveis.

1. (3,0) Vamos procurar sarna pra se coçar? (esse exercício tem por objetivo te ensinar procurar sarna para se coçar!! – entre base de dados na internet e crie problemas como este que inventei para você)

**Consumo de Combustível**

* + **Descrição**: Utilize dados sobre o consumo de combustível de veículos, incluindo variáveis como cilindradas, peso do veículo, tipo de combustível, etc.
  + **Tarefa**: Crie um modelo de regressão linear para prever o consumo de combustível. Avalie a heterocedasticidade e, se necessário, aplique uma transformação Box-Cox para melhorar o modelo.
  + **Desafios**: Base de dados reais devem ser bem analisadas porque sempre tem merda, né? Para cada problema que encontrar... treine fazer perguntas (lembre-se de documentar), e suponha respostas (documente também, para seu trabalho fazer sentido) para que você possa seguir com o trabalho (porque você já está careca de saber que não se deve supor nada, NUNCA).
  + **Fonte de Dados**: UCI Machine Learning Repository - Auto MPG Data

<https://archive.ics.uci.edu/ml/datasets/auto+mpg>

Arquivo (Regressão Linear BoxCox.PDF) anexo com os código, análises e conclusões.

Gerei um modelo usando Step Wise e transformação do Y usando BoxCox. Para verificar o y é preciso reverter a transformação BoxCox que foi aplicada na variável Y (alvo).

O modelo de regressão com uso de BoxCox apresentou resulatados não muito bons (satisfatórios) e com isso, gerei um outro modelo, porém usando PCA, que ficou absurdamente melhor, porém nada fácil de explicar o comportamento.