

## 4ª Lista de exercícios utilizando o R/Rmarkdown

A base de dados "**Base\_DEA\_valores\_medios\_2014-2016.csv**" contém os valores médios, considerando o período 2014 a 2016 das seguintes variáveis:

- Custos Operacionais (PMSO)
- Rede subterrânea (Km)
- Rede de distribuição aérea (Km)
- Rede de alta tensão (Km)
- Mercado ponderado (MWh)
- Consumidores totais (unid)
- Perdas Não Técnicas (MWh)
- Consumidor Hora Interrompido - CHI (h)

Essas variáveis estão disponíveis para 61 empresas distribuidoras de energia elétrica. A Agência Nacional de Energia Elétrica (ANEEL) utilizou essas informações para calcular a eficiência operacional das empresas de distribuição no ano de 2017.

Neste exercício, o objetivo é avaliar se os Custos Operacionais (PMSO) podem ser preditos a partir das demais variáveis. Serão analisados dois modelos:

1. Procure ajustar um modelo de regressão linear múltiplo onde a variáveis resposta é o PMSO (y). Procure identificar o modelo que seja estatisticamente factível.
2. Procure ajustar um modelo de regressão linear múltiplo utilizando TODAS as variáveis disponíveis para predizer o PMSO. Procure ajustar o modelo de forma que o mesmo **seja o melhor modelo preditivo**.
3. Procure incorporar na sua análise o modelo **CART** (pacote rpart e rpart.plot) realizando a análise de um único modelo **CART** e de um modelo híbrido combinando o modelo linear múltiplo e o modelo **CART**.

É conhecido que os dados apresentam heterocedasticidade. Uma solução é aplicar o logaritmo na variáveis resposta (PMSO) e nas variáveis preditoras, quando pertinente.

Faça uma comparação dos modelos ajustados.

Elabore o seu relatório utilizando o **Rmarkdown** e envie o documento em formato **PDF** para avaliação pelo sistema minha.ufmg.