



PONTIFÍCIA UNIVERSIDADE CATÓLICA DE GOIÁS (PUC GOIÁS) ESCOLA
POLITÉCNICA – MESTRADO EM ENGENHARIA DE PRODUÇÃO E
SISTEMAS (MEPROS)

1º DESAFIO EM CIÊNCIA DE DADOS

Roteiro para análise dos dados:

Objetivo 1: Fazer um tratamento no conjunto de dados (BASE_DESAFIO.xlsx)

1. Exame gráfico dos dados (exames de relações entre variáveis, exames das diferenças de grupos, identificações de observações atípicas)
2. Cálculo de medidas resumos para o conjunto de dados (medidas de posição, dispersão e de forma)
3. Nas variáveis métricas fazer as verificações de Normalidade e ações corretivas para não normalidade.
4. Identificar relações de dependências entre as variáveis (homoscedasticidade e Linearidade) e transformações para atingir normalidade, homocedasticidade e linearidade.

Objetivo 2: Reduzir as percepções da EMPRESA_DESAFIO sobre as 13 variáveis métricas (X6 a X18) a um menor número de variáveis. Definir a técnica a ser utilizada e descrever o resultado.

Objetivo 3: A administração da EMPRESA_DESAFIO há muito tem se interessado por previsões mais precisas do nível de satisfação de seus clientes. Se bem-sucedidas, isso fornece uma melhor fundamentação para seus esforços de *marketing*. Para este propósito, selecione a variável satisfação do cliente (X19) como variável dependente (Y) a ser prevista por variáveis independentes que representam percepções da atuação da EMPRESA_DESAFIO pelos seus clientes (variáveis: X6 a X18). Escolha um modelo que melhor se ajusta ao problema. Se for necessário pode-se descartar algumas das variáveis independentes ou mesmo utilizar as variáveis obtidas no objetivo 2, nessa modelagem.

Objetivo 4: Fazer estimação do modelo encontrado no objetivo anterior (objetivo 3) e avaliar a qualidade do ajuste do modelo encontrado.

Objetivo 5: Avaliar se existe diferença de percepções da EMPRESA_DESAFIO (X6 a X18) entre os clientes do EUA/América do Norte e aqueles do resto do mundo (X4). Sugestão: Use um modelo de regressão logística para responder o problema.