



LABORATÓRIO 7

Gastos com Saúde nos EUA

Retomando os dados do laboratório 2, existe uma série de possibilidades a serem exploradas. Buscando facilitar a análise, os dados foram filtrados de modo que foram considerados apenas pacientes fumantes e que moram na região sudeste dos Estados Unidos. A partir desses dados, podemos conduzir uma série de análises considerando um modelo de regressão linear múltipla entre o custo do seguro saúde, o IMC dos pacientes e a idade. Os dados estão disponíveis no arquivo “lab7.csv”.

Objetivos

1. Faça uma análise descritiva considerando gráficos uni e bivariados, medidas de posição e dispersão e correlação.
2. Ajuste um modelo de regressão linear múltipla entre a variável charges e as demais variáveis explicativas quantitativas.
3. Escreva a equação do modelo ajustado e interprete os seus coeficientes.
4. Conduza testes para determinar quais variáveis são estatisticamente significantes ao nível de significância de 5%.
5. Obtenha o quadro da análise de variância e interprete o resultado do teste F para avaliar a bondade do ajuste do modelo.
6. Obtenha o coeficiente de determinação e o coeficiente de determinação ajustado do modelo e interprete-os.
7. Faça os gráficos de diagnóstico e comente-os
 - (a) Valores Ajustados e Resíduos Studentizado
 - (b) Gráfico Quantil-Quantil
 - (c) Gráfico de Distância de Cook
 - (d) Gráfico dos pontos de Alavanca e Resíduo Studentizado
 - (e) Gráfico de DfBeta
 - (f) Gráfico de DfFit
 - (g) Gráfico do COVRatio

Descrição das Variáveis

- *age*: idade em anos completos do paciente
- *charges*: preço do seguro saúde em dólares
- *bmi*: IMC do paciente

Dicas:

- Intervalos de confiança: comandos
`predict(nome_modelo, data.frame(X1=20, X2=20), interval="confidence", level=0.90)`
`predict(nome_modelo, data.frame(X1=20, X2=20), interval="prediction", level=0.90)`
- Resíduos Studentizados: comando `rstudent(nome_modelo)`
- *Leverages*: comando `hat(x)` em que `x <- model.matrix(nome_modelo)`. Os pontos de alavanca devem ser identificados (comando `identify`).
- Distâncias de Cook: comando `cooks.distance`.
- Medidas de influência: comandos `influence.measures`, `dfbetas`, `dffits` e `covratio`.
- Colinearidade: comando `vif` do pacote `faraway`.