

Testar a multicolinearidade

O modelo de regressão linear pressupõe que as variáveis explicativas não estão correlacionadas. Pode-se detectar a multicolinearidade, isto é, correlação existente entre variáveis independentes:

- Grandes modificações nas estimativas dos coeficientes de regressão quando uma variável independente é adicionada ou removida do modelo.
- Elevadas variâncias (t-values pequenos) dos coeficientes de regressão de variáveis independentes que se esperavam ser importantes no modelo.
- Elevados coeficientes de correlação entre os pares de variáveis independentes indicam uma forte relação linear entre elas.
- A relação entre variáveis independentes pode ser avaliada usando:
 - a matriz de correlação das variáveis independentes. Se existirem pares de variáveis altamente correlacionadas, então deve-se eliminar uma delas, deixando aquela que fizer mais sentido do ponto de vista prático.
 - VIF(variance inflation factor):

$$VIF = \frac{1}{1 - R_j^2}$$

onde R_j^2 é o coeficiente de determinação da regressão da variável x_j em função das outras variáveis independentes. Quando $VIF \approx 1$, temos que as variáveis independentes são independentes e quando VIF é maior que 10 implica que as variáveis independentes estão linearmente dependentes (este ponto de corte é arbitrário)