

Modelos Lineares

Regressão Linear Múltipla

Susana Faria

Relembremos que de acordo com o modelo de regressão linear múltipla erros das observações satisfazem os seguintes pressupostos:

- seguem uma distribuição normal;
 - têm media zero;
 - têm variância constante (homocedasticidade);
 - são independentes.
-
- A verificação das hipóteses é fundamental, visto que toda a inferência estatística no modelo de regressão linear (testes de hipóteses) se baseia nesses pressupostos.
 - Nesse sentido, se houver violação dos mesmos, a utilização do modelo deve ser posta em causa.
 - A **Análise dos Resíduos** é uma ferramenta usada para detectar violações dos pressupostos.
 - Os métodos para validar o modelo são basicamente os mesmos apresentados para a regressão simples.

Normalidade dos erros

- O pressuposto da normalidade pode ser testado traçando um Normal QQ-plot ou um Normal PP-plot para os resíduos. **Se os erros possuírem distribuição Normal, todos os pontos dos gráficos devem posicionarem-se mais ou menos sobre uma recta.**
- Também se pode proceder a testes de ajustamento dos resíduos a uma distribuição Normal: Teste Kolmogorov- Smirnov e Teste de Lilliefors.

Média Nula, Variância constante e independência dos erros

- Estes pressupostos podem ser verificados graficamente representando os resíduos versus valores estimados da variável dependente \hat{Y}_i (ou versus valores da variável independente).
- **Os pontos do gráfico devem distribuir-se de forma aleatória em torno da recta que corresponde ao resíduo zero, formando uma mancha de largura uniforme.** Dessa forma será de esperar que os erros sejam independentes, de média nula e de variância constante.
- Quando os resíduos não se comportam de forma aleatória, ou seja, seguem um padrão, a condição de independência não é satisfeita.
- Se a dispersão dos resíduos aumentar ou diminuir com os valores da variável independentes x_i , ou com os valores estimados da variável dependente \hat{Y}_i , deve ser posta em causa a hipótese de variâncias constante dos erros.

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)