

Profissão: Cientista de Dados



GLOSSÁRIO



Regressão III



Dica: para encontrar rapidamente a palavra que procura aperte o comando CTRL+F e digite o termo que deseja achar.

- **Realize inferência em regressão**
- **Faça inferência sobre Y**
- **Transforme em X**
- **Aplique regressão segmentada**
- **Utilize Lowess**



Realize inferência em regressão



Realize inferência em regressão

● Intercepto

O valor da variável dependente quando todas as variáveis independentes são zero.

● Intervalo de confiança

Um intervalo de valores, derivado dos dados da amostra, que é provável que contenha o valor verdadeiro de um parâmetro desconhecido.



Faça inferência sobre Y



Faça inferência sobre Y

• LOESS

Locally Estimated Scatterplot Smoothing.
Técnica mais geral que tenta encontrar
uma curva suave que se ajuste aos
dados.

• pd.qcut

Função do pacote pandas em Python usada
para dividir a variável explicativa em
categorias.



Transforme em X



Transforme em X

● Relações Não Lineares

Relações entre variáveis que não podem ser representadas por uma linha reta.

● Transformação Logarítmica

Uma transformação que aplica o logaritmo aos dados.



Aplique regressão segmentada



Aplique regressão segmentada

● Constante C_0

É o ponto em que a inclinação da reta de regressão muda na regressão segmentada.

● Regressão Segmentada

É um método que mantém a interpretabilidade do modelo, mantém a parcimônia e oferece mais flexibilidade a um modelo de regressão. A ideia é modelar uma reta de regressão até um certo ponto (C_0) e, a partir desse ponto, a reta muda de inclinação.



Utilize Lowess



Utilize Lowess

• Função de Interpolação

Uma função de interpolação é uma função que é usada para estimar os valores de uma variável dependente a partir dos valores de uma ou mais variáveis independentes.

• Lowess (Locally Weighted Scatterplot Smoothing)

Lowess é uma técnica de suavização de dados que é usada para estimar o valor de uma variável dependente a partir dos valores de uma variável independente. Ela dá mais peso aos pontos de dados que estão próximos ao ponto de interesse e menos peso aos pontos de dados que estão mais distantes.



Bons estudos!

