

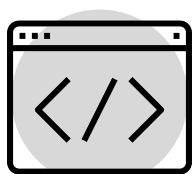
# Profissão: Cientista de Dados



# BOAS PRÁTICAS



# Regressão III



- **Faça inferência sobre Y**
- **Categorização para linearização**
- **Transforme em X**
- **Aplique regressão segmentada**
- **Utilize Lowess**



# Faça inferência sobre Y

- Sempre considere a incerteza ao fazer previsões: Ao fazer previsões usando regressão, é importante considerar a incerteza tanto nos parâmetros do modelo (alfa e beta) quanto na variância da variável dependente Y.
- Lembre-se de que o intervalo de confiança para a média é mais estreito do que para uma nova observação: Isso reflete as incertezas adicionais na estimativa de uma nova observação.
- Use a visualização para entender suas previsões: Plotar suas previsões e intervalos de confiança pode ajudar a entender melhor o desempenho do seu modelo.
- Esteja ciente de que o modelo de mínimos quadrados ordinários pode não se ajustar perfeitamente aos dados: Isso pode ser abordado em análises futuras e é importante estar ciente disso ao interpretar os resultados.



# Categorização para linearização

- Ao lidar com relações não lineares em um modelo de regressão linear, considere técnicas como a categorização da variável explicativa, aplicação de uma função, regressão segmentada e LOESS. A escolha da técnica depende do conjunto de dados específico e do objetivo da análise.
- Ao aplicar uma função à variável explicativa, lembre-se de que isso pode ajudar a ajustar o modelo a um padrão não linear nos dados, mas pode ser difícil encontrar a função correta.
- Ao categorizar a variável explicativa, divida a variável em categorias lógicas e modele cada categoria separadamente. Isso pode resultar em um modelo que se parece com uma "função escada", onde cada degrau representa uma categoria diferente.
- Ao aplicar uma função à variável explicativa, lembre-se de que isso pode ajudar a ajustar o modelo a um padrão não linear nos dados, mas pode ser difícil encontrar a função correta.



# Transforme em X

- Experimente diferentes transformações: Não se limite a uma única transformação. Experimente diferentes transformações, como logarítmica, quadrática, senoidal, etc., para ver qual delas fornece o melhor ajuste.
- O valor da constante adicionada pode afetar a curvatura do ajuste: Ao aplicar uma transformação logarítmica, o valor da constante adicionada pode afetar a curvatura do ajuste. Experimente diferentes valores para encontrar o melhor ajuste.
- Lembre-se de que o logaritmo de zero é indefinido: Ao aplicar uma transformação logarítmica, lembre-se de que o logaritmo de zero é indefinido. Para evitar esse problema, adicione uma constante aos dados.
- O objetivo é encontrar um modelo que se ajuste bem aos dados, e não o contrário: Não force os dados a se ajustarem a um modelo específico. Em vez disso, procure um modelo que se ajuste bem aos dados.





# Aplique regressão segmentada

- Ao usar a regressão segmentada, é importante garantir que o modelo ainda seja interpretável. Isso significa que você deve ser capaz de explicar como as variáveis independentes estão relacionadas à variável dependente.
- Ao implementar a regressão segmentada, é importante definir o ponto ( $C_0$ ) em que a inclinação da reta de regressão muda. Isso pode ser feito através de análise exploratória dos dados ou com base em conhecimento de domínio.
- Para implementar a regressão segmentada, você precisará criar duas novas variáveis,  $X_a$  e  $X_b$ . Essas variáveis ajudarão a definir o ponto em que a inclinação da reta de regressão muda.
- Lidar com outliers ou erros de entrada de dados: Na regressão segmentada, outliers ou erros de entrada de dados podem afetar o modelo. Uma boa prática é truncar a variável  $X$  em um valor de corte para evitar danos ao modelo.



# Utilize Lowess

- 
 Ao usar a técnica Lowess, é importante entender que ela dá mais peso às variáveis próximas ao X em questão e menos peso às variáveis mais distantes. Portanto, é uma ferramenta útil para estimar o valor de Y em função de X.
- 
 Ao usar a técnica Lowess, é importante lembrar que ela não está disponível no pacote statsmodels, mas pode ser encontrada no pacote scipy. Portanto, é necessário ter familiaridade com diferentes pacotes e saber onde encontrar as funções necessárias.





# Utilize Lowess

- A técnica Lowess não fornece uma interpretação tão direta e concisa quanto a regressão, mas fornece uma curva suave que passa aproximadamente pela nuvem de pontos dos dados. Portanto, é importante ser capaz de interpretar os resultados do Lowess e entender o que a curva está mostrando.
- Ao ajustar um modelo usando a função Lowess, é necessário definir  $Y$ ,  $X$  e um hiperparâmetro que determina a fração da base de dados que será usada para estimar  $Y$ . Portanto, é importante entender como os hiperparâmetros afetam o modelo e como ajustá-los adequadamente.



# Bons estudos!

