

Profissão: Cientista de Dados



GLOSSÁRIO



Regressão IV



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

- **Apresente a Introdução**
- **Avalie as Suposições**
- **Identifique Outliers**
- **Aborde Casos de Correções**
- **Analise a Correlação nas Variáveis Explicativas – Multicolinearidade**



Apresente a Introdução



Apresente a Introdução

• Aplicação Cruzada

O uso de um modelo para um propósito para o qual ele não foi originalmente projetado. Isso pode levar ao risco de modelo, pois o modelo pode não funcionar tão bem em um contexto diferente.

• Entrada de informações

Um dos três componentes de um modelo, conforme definido pelo Fed. Refere-se aos dados que são inseridos no modelo para processamento.

• Limitações do modelo

Refere-se às restrições e falhas potenciais de um modelo de regressão. É importante entender essas limitações para evitar a aplicação inadequada do modelo.



Avalie as Suposições



Avalie as Suposições

Heterocedasticidade

Refere-se à situação em que a variância dos erros ou resíduos não é constante em todos os níveis de uma variável independente. No exemplo da aula, a variância parece aumentar com o valor da conta, indicando heterocedasticidade. Se a suposição de homocedasticidade falhar, pode comprometer as estimativas do modelo.

Homocedasticidade

É uma das suposições em um modelo de regressão. Significa que a variância dos erros ou resíduos é a mesma em todos os níveis de uma variável independente. É o oposto da heterocedasticidade.



Avalie as Suposições

Independência dos resíduos

É uma suposição em um modelo de regressão que indica que os erros ou resíduos são independentes uns dos outros. Ou seja, o erro em uma observação não afeta o erro em outra observação.

Resíduos

Em um modelo de regressão, os resíduos são a diferença entre os valores observados da variável dependente e os valores previstos pelo modelo. As suposições de um modelo de regressão são baseadas nos resíduos.



Identifique Outliers



Identifique Outliers

• Distorção

É o efeito que os outliers podem ter em um modelo de regressão, potencialmente enviesando os resultados e afetando a precisão do modelo.

• Resíduo

É a diferença entre o valor observado e o valor previsto por um modelo de regressão. O valor do resíduo pode ser usado para identificar possíveis outliers.

• Ponto de Alavanca

É um ponto de dados que tem uma grande influência na linha de regressão. Um ponto de alavanca pode alterar significativamente a linha de regressão.

• Transformação

É uma maneira de lidar com outliers. Pode envolver a alteração dos dados de uma maneira que minimize o impacto dos outliers.



Identifique Outliers

Trucagem

É outra maneira de lidar com outliers. Envolve a remoção de outliers do conjunto de dados.



Aborde Casos de Correções



Aborde Casos de Correções

Independência das variáveis explicativas

Refere-se à ausência de qualquer relação entre as variáveis independentes em um modelo de regressão. Se as variáveis independentes são independentes, isso significa que a mudança em uma variável não afeta as outras.

Multilinearidade

É um termo usado para indicar que há uma relação linear entre duas ou mais variáveis independentes.

Logaritmo (log)

É uma função matemática que é usada para transformar uma série de dados. A transformação logarítmica é uma maneira comum de lidar com dados distorcidos.

Pontos de dados influentes

São pontos de dados que têm um grande efeito sobre a estimativa dos parâmetros de um modelo de regressão.



Aborde Casos de Correções

Resíduo estudante

É uma medida de quão bem um ponto de dados se ajusta a um modelo estatístico. É calculado como a diferença entre o valor observado e o valor previsto pelo modelo, dividido pelo erro padrão da estimativa.



Analise a Correlação nas Variáveis Explicativas – Multicolinearidade



Analise a Correlação nas Variáveis Explicativas – Multicolinearidade

Coeficiente V de Cramer

Medida de associação entre duas variáveis qualitativas. Varia de 0 (nenhuma associação) a 1 (associação perfeita).

Correlação de Pearson

Medida da relação linear entre duas variáveis contínuas. Varia de -1 (correlação negativa perfeita) a 1 (correlação positiva perfeita).

Correlação de Spearman

Medida da relação monotônica entre duas variáveis. Varia de -1 (correlação negativa perfeita) a 1 (correlação positiva perfeita).

Multicolinearidade

Situação em que duas ou mais variáveis explicativas em um modelo de regressão estão altamente correlacionadas, levando a resultados instáveis.



Analise a Correlação nas Variáveis Explicativas – Multicolinearidade

Remoção do efeito de uma variável em outras variáveis

Método para lidar com a multicolinearidade que envolve ajustar um modelo de regressão para a variável de interesse e as variáveis que estão causando a multicolinearidade, e então usar os resíduos desse modelo como a nova variável de interesse.

Variance Inflation Factor (VIF)

Medida que indica o quanto a variância dos coeficientes de regressão estimados é aumentada devido à multicolinearidade. Um VIF alto indica alta multicolinearidade.



Bons estudos!

