



ANALISTA DE DADOS

| Nov22

REGRESSÃO

Análise de Regressão

❑ OBJETIVO:

Estudar o relacionamento entre uma variável dependente (alvo/resposta) e uma ou mais variáveis independente (preditor).

Ajuda a entender como o valor da variável dependente está mudando quando levamos em consideração uma variável independente enquanto outras variáveis independentes são mantidas fixas.

❑ PROCEDIMENTO:

Obter uma relação funcional (modelo matemático) que associe a variável resposta com as variáveis preditoras (uma ou mais). Ou seja, partir do conhecimento do valor das variáveis preditoras, possamos obter o máximo possível de informações sobre a variável resposta, isto é: $Y = f(x)$.

ANÁLISE DE REGRESSÃO

Alguns usos

Usado principalmente para predição, previsão, modelagem de série temporal e determinação da relação causal-efeito entre as variáveis.

- Explicar situações que se quer entender: Por que as ligações para o atendimento ao cliente caíram no mês passado?
- Prever algo sobre o futuro: Como serão as vendas nos próximos seis meses?
- O risco de uma doença cardíaca, dado o histórico familiar e uma série de características físicas e psicológicas.
- Em finanças, calcular o Beta (volatilidade dos retornos em relação ao mercado geral) para uma ação.



ANÁLISE DE REGRESSÃO

Modelos de Regressão

QUESTÕES:

- Se LINEAR, qual o número de variáveis respostas?

- UMA \rightarrow Univariada

$$Y_{n \times 1} = f(x)$$

Ex: A partir do valor do Peso, prever valor da Altura

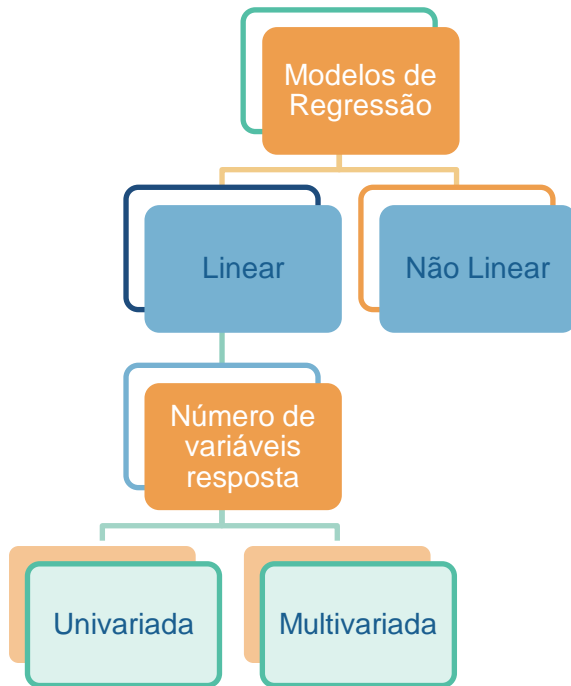
- DUAS OU MAIS \rightarrow Multivariada

$$Y_{n \times p} = f(x)$$

- **Ex:** A partir do valor do Peso, prever valor da Altura e da Circunferência da cintura.

ANÁLISE DE REGRESSÃO

Modelos de Regressão



ANÁLISE DE REGRESSÃO

Modelos de Regressão

QUESTÕES:

- Se UNIVARIADO, qual o número de variáveis preditoras?

- UMA → **Simples**

$$Y_{nx1} = f(x_{nx2})$$

Ex: A partir do valor do Peso, prever valor da Altura

- DUAS OU MAIS → **Múltipla**

$$Y_{nx1} = f(x_{nxp}), p > 2$$

- **Ex:** A partir do valor do Peso, da Circunferência da cintura, do Gênero e da Idade, prever o valor da Altura.

ANÁLISE DE REGRESSÃO

Modelos de Regressão



ENTENDENDO O PROBLEMA

ANÁLISE DE REGRESSÃO

Entendendo o Problema

O banco de dados a ser analisado é composto por 5 variáveis:

- Ano;
 - Mês;
 - Taxa de juros;
 - Taxa de desemprego;
 - Preço de ações.
- a) Quais seriam as variáveis preditoras?
 - b) Qual seria a variável de interesse?
 - c) Por que deveríamos usar regressão?



ANÁLISE PRELIMINAR

ANÁLISE DE REGRESSÃO

Análise Preliminar

Algumas análises que podemos fazer nesta etapa são:

- Calcular as medidas-resumo e visualizá-las em um *boxplot* para detectar possíveis outliers;
- Verificar se alguma variável categórica/binária precisa de algum tratamento;
- Utilizar gráficos de dispersão para verificar quais as melhores “candidatas” para integrar o modelo.



EXECUÇÃO DO MODELO

ANÁLISE DE REGRESSÃO

Execução do Modelo

Interpretações que necessitam ser realizadas nesta etapa:

- Na seção de resíduos, verificar se a mediana está próxima de zero e se os valores máximo e mínimo são iguais (ou próximos) em valores absoluto;
- Interpretar os coeficientes do modelo;
- Analisar a significância de cada variável no modelo pelo seu p-valor;
- Calcular intervalos de confiança para seus coeficientes.



VALIDAÇÃO DO MODELO

REGRESSÃO

Métricas de qualidade do ajuste

Para verificar a qualidade do ajuste de um modelo, utiliza-se **três métricas**:

- RSE (Erro padrão dos Resíduos);
- R-Quadrado (R^2);
- Estatística-F.



ANÁLISE DE REGRESSÃO

Métricas de qualidade do ajuste - RSE

O RSE (Erro Padrão dos Resíduos) mede o desvio dos resíduos, ou seja, a influência daquilo que o modelo não consegue controlar ou captar.

Queremos que o RSE seja o menor possível, pois ele representa os padrões que não foram explicados pelo modelo.

$$\sqrt{\frac{1}{n-2} \sum_{i=1}^n (y_i - \hat{y}_i)^2}$$

ANÁLISE DE REGRESSÃO

Métricas de qualidade do ajuste - RSE

O RSE (Erro Padrão dos Resíduos) mede o desvio dos resíduos, ou seja, a influência daquilo que o modelo não consegue controlar ou captar.

Queremos que o RSE seja o menor possível, pois ele representa os padrões que não foram explicados pelo modelo.

$$\sqrt{\frac{1}{n-2} \sum_{i=1}^n (y_i - \hat{y}_i)^2}$$



ANÁLISE DE REGRESSÃO

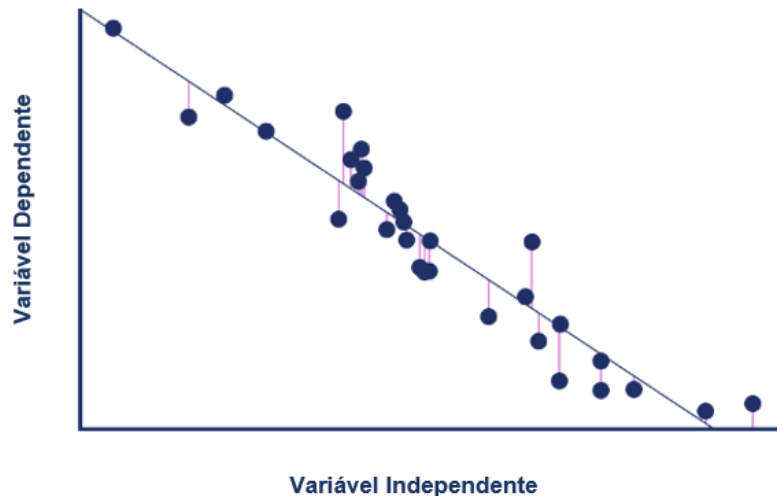
Métricas de qualidade do ajuste – R^2

- O coeficiente de determinação (R^2) mede quão bem um modelo estatístico prevê um resultado (variável dependente do modelo).
- O **menor valor possível** de R^2 é 0 e o **maior valor possível** é 1. Ou seja, quanto melhor um modelo for em fazer previsões, mais próximo seu R^2 estará de 1.
- Mais tecnicamente, R^2 é uma medida de quão bom está o ajuste pois é a proporção da variância na variável dependente que é explicada pelo modelo.

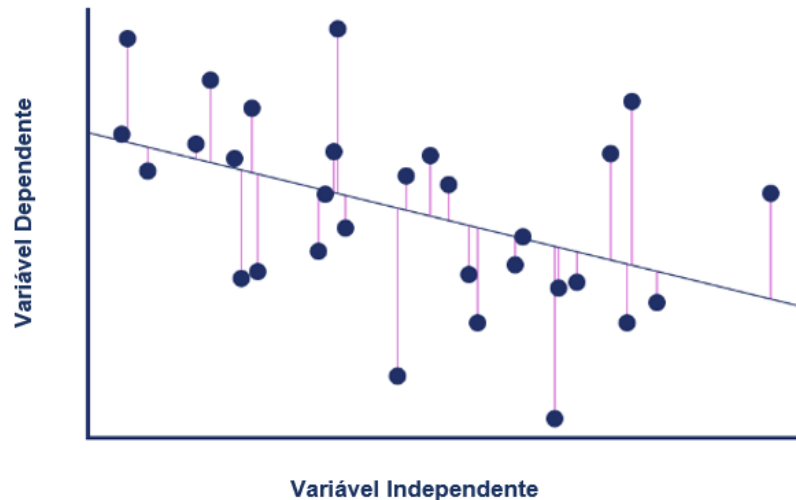
ANÁLISE DE REGRESSÃO

Avaliação do modelo – R^2

Coeficiente de determinação (R^2) = 0.9



Coeficiente de determinação (R^2) = 0.2



LIMITAÇÕES

- Não performa bem em modelos não lineares e categóricos;
- Não determina se as estimativas e previsões dos coeficientes são tendenciosas/enviesadas.

PREDIÇÕES