

---

---

# Machine Learning

19/2/2024  
Me. Fernando Nemec

---

---

# Regressão Linear

Regressões Lineares são uma das formas mais simples e utilizadas para prever resultados baseados em um amplo conjunto de dados.

A técnica simplesmente se refere a produzir uma função de primeiro grau (reta) que passe se aproxime da melhor forma possível dos valores utilizados para treinar o modelo.

# Equação

A equação de uma Regressão Linear se assemelha a isto:

$$Y = \beta_0 + \beta_1 X + e$$

Variável  
reposta

Intecepto

Coefficiente  
angular

Variável  
explicativa

Erro

# Equação

$Y$  ou  $\hat{Y}$  é a variável **regredida**, também chamada de variável de interesse, ou seja, aquela que pretendemos obter após rodar o nosso modelo.

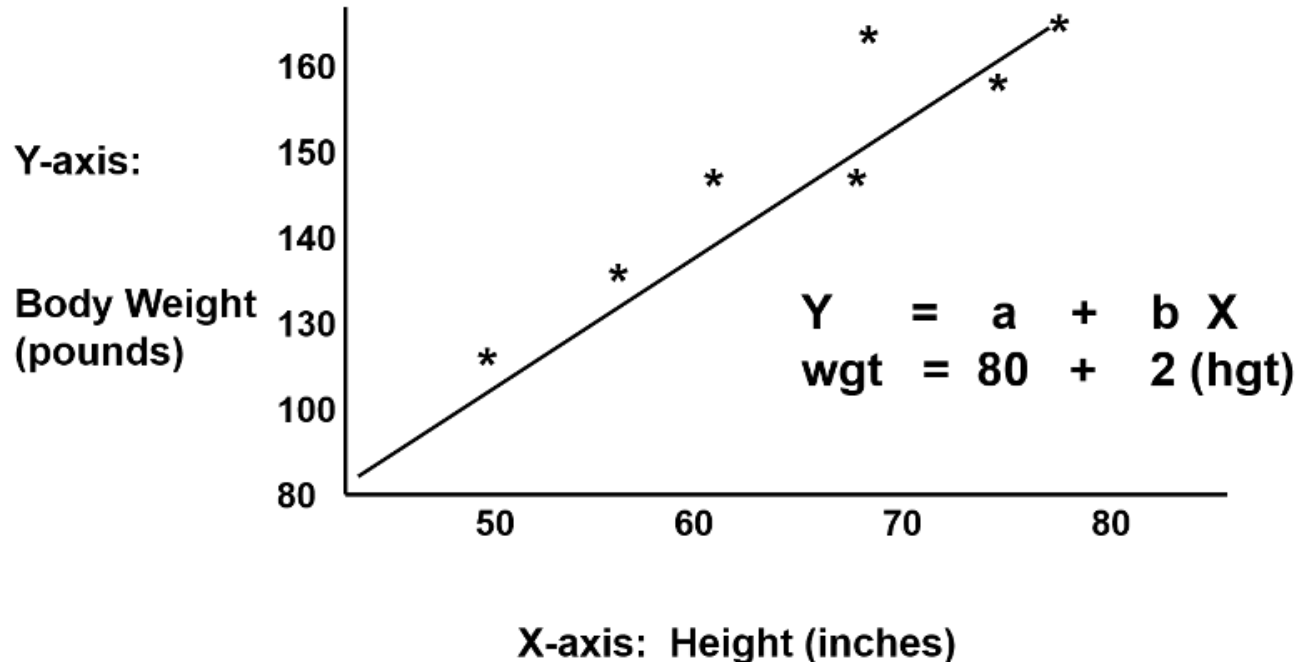
$\beta_0$  é o nosso intercepto ou a constante da nossa equação de primeiro grau. Ela indica onde a nossa reta intercepta o eixo  $y$ .

$\beta_1$ , ou coeficiente angular, indica a inclinação da nossa reta que é calculada de forma a se aproximar da melhor forma de todos os nossos pontos de interesse.

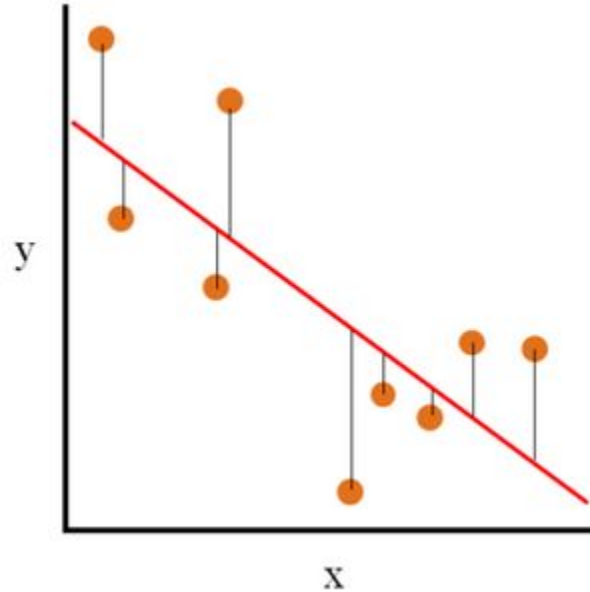
$X$  é o **regressor ou preditor**. Também é chamada de variável explicativa. É a variável que vamos utilizar para predizer o valor da nossa variável de interesse.

$E$  é o termo de erro e indica tudo aquilo que não enxergamos ao criar o nosso modelo.

# Gráfico de uma Regressão Linear Simples



# Gráfico de uma Regressão Linear Simples



# Correlação de Pearson

Uma forma de determinar se um regressor consegue prever com eficiência uma variável explicativa é através do coeficiente de correlação.

Um coeficiente próximo de 1 ou -1 sugere que a variável será um bom regressor.

# Erro de uma Regressão

Uma regressão sempre terá erros pois dificilmente a equação obtida passará sobre todos os pontos de forma perfeita. Isso indica que para cada previsão do treino há um pequeno erro. O R-quadrado utiliza esses erros para expressar o poder de explicação de um modelo.

Quanto mais próximo de 1, melhor o poder de explicação de um modelo. Um valor próximo de 0, indica que o modelo não explica absolutamente nada sobre os dados.



# Prática

Pacotes necessários:

- numpy
- pandas
- matplotlib
- sklearn.linear\_model
- seaborn
- statsmodels