

Regressão Linear Simples

Equação e proposta

Objetivo: prever um valor chave, nesse caso nossa variável X_1 . A equação é descrita por: $\hat{y} = b_0 + b_1 X_1$, em que:

- \hat{y} é a variável que queremos prever
- b_0 é uma constante
- X_1 é o preditor
- b_1 é o coeficiente de inclinação

Em um caso hipotético, podemos definir:

$$Potatoes(t) = b_0 + b_1 \times Fertilizer (kg)$$

- $b_0 = 8(t)$
- $b_1 = 3\left(\frac{t}{kg}\right)$

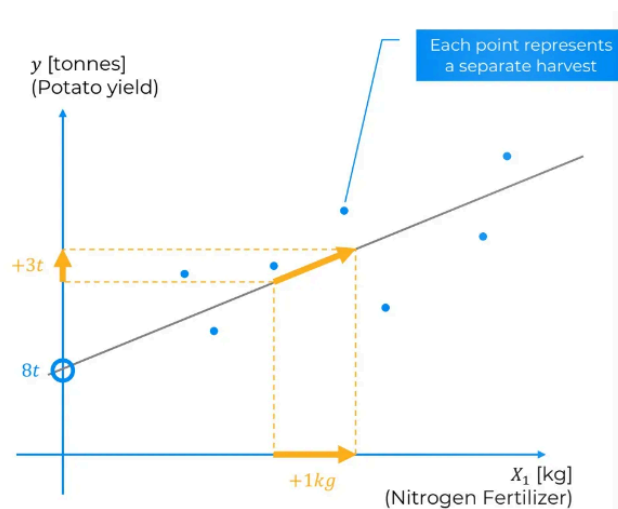


Figura 1: Scatterplot com a linha predita.

A **Figura 1** os pontos em azul representam o plot das colheitas, no eixo X (quantidade de fertilizante) e no eixo Y a quantidade em toneladas da quantidade de batatas colhidas. A linha em preto representa a Regressão Linear Aplicada aos dados, que tem por objetivo traçar a melhor reta para representar um caso não observado da colheita.

Método dos Mínimos Quadrados

Ao gerarmos uma regressão linear, criamos várias linhas que podem representar o nosso conjunto de dados, qual delas é a que melhor representa a realidade?

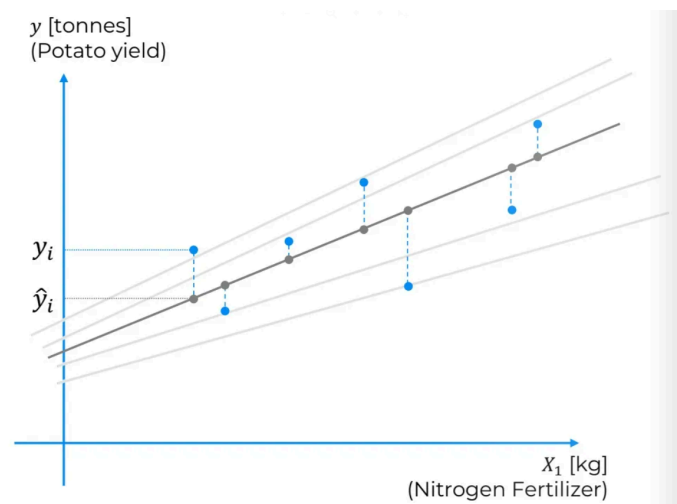


Figura 2: Representação de todas as linhas geradas.

Para responder isso, utilizamos o conceito de resíduo: $\epsilon_i = y_i - \hat{y}_i$ em que y_i é a observação e \hat{y}_i é o valor da regressão observado para um mesmo X_i . O objetivo final, portanto, é encontrar uma reta em que $\sum (y_i - \hat{y}_i)$ é minimizado.