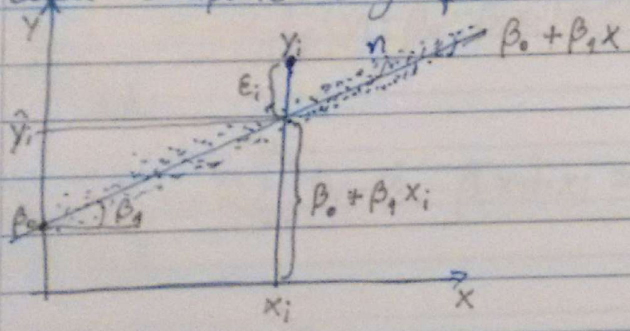
Entrega 2: Isabella e Patrick dos Santos

· **Objetivo:** Aprimoramento do nosso conhecimento da técnica estatística regressão a fim de estudarmos a relação matemática existente entre nossas variáveis de interesse.

· **Desenvolvimento:**

→ Obter os estimadores $\hat{\beta}_0$ e $\hat{\beta}_1$ (estimativa dos valores mínimos do Erro total):

Para achar esses estimadores, é preciso entender os componentes de um gráfico da relação linear de duas variáveis: Assim para facilitar o entendimento iremos explicar cada componente gráfico:



y_i = valor da variável resposta
x_i = valor da variável explicativa
β_0 = parâmetro do intercepto
β_1 = parâmetro do coeficiente angular
$\hat{y}_i$ = regressão de Y em X
ε_i = erro estocástico
$i = 1, 2, 3, \ldots, n$.
n = Tamanho da amostra

Com isso podemos começar a estimar os valores de $\hat{\beta}_0$ e $\hat{\beta}_1$, a partir de uma amostra de n pares de valores (X_i, Y_i) com i igual à $1, \ldots, n$, que são os n pontos de um gráfico. Supondo que é traçada uma reta que passa o mais próximo possível de todos os pontos, essa reta será definida por $\beta_0 + \beta_1 X$. Existe também uma diferença entre os pontos e a reta, que é chamado de desvio ou erro estocástico. Esse erro pode ser definido por $e_i = Y_i - (\beta_0 + \beta_1 X_i)$.

Aqui visamos obter valores estimados de β_0 e β_1, sempre buscando valores mínimos para o desvio e_i. Uma técnica muito eficiente para este cálculo é o Método do Mínimos Quadrados. Nele, minimizamos a soma dos quadrados mínimos na seguinte expressão:

$$L=\sum_{i=1}^{n} e_i^2=\sum_{i=1}^{n}(y_i-\beta_0-\beta_1 x_i)^2$$

a expressão é elevado ao quadrado para que os valores seja sempre positivos

As estimativas para os parâmetros β_0 e β_1 precisam satisfazer as expressões derivativas:

$$\left.\frac{\partial L}{\partial \beta_0}\right|_{\hat{\beta}_0,\hat{\beta}_1}=-2\sum_{i=1}^{n}(y_i-\hat{\beta}_0-\hat{\beta}_i x_i)=0$$

$$\left.\frac{\partial L}{\partial \beta_1}\right|_{\hat{\beta}_0,\hat{\beta}_1}=-2\sum_{i=1}^{n}(y_i-\hat{\beta}_0-\hat{\beta}_i x_i)\cdot x_i=0$$

Igualamos à 0, pois substituímos por $\hat{\beta}_0$ e $\hat{\beta}_1$, que são valores particulares de L

Após simplificar essas equações, iremos chegar às novas equações que são chamadas de Equações Normais de Mínimos Quadrados

$$\begin{cases} n\hat{\beta}_0+\hat{\beta}_1\sum_{i=1}^{n}x_i=\sum_{i=1}^{n}y_i \\ \hat{\beta}_0\sum_{i=1}^{n}x_i+\hat{\beta}_1\sum_{i=1}^{n}x_i^2=\sum_{i=1}^{n}x_i y_i \end{cases}$$

Equações Normais de Mínimos Quadrados

Resolvendo o sistema de equações descobrimos que:

$$\hat{\beta}_0=\frac{\sum_{i=1}^{n}y_i}{n}-\frac{\hat{\beta}_1\sum_{i=1}^{n}x_i}{n}=\bar{Y}-\hat{\beta}_1\bar{X}$$

Scanned by CamScanner

Substituindo o valor de $\hat{\beta}_0 (\bar{Y} - \hat{\beta}_1 \bar{X})$ na segunda equação da Equações Normais de Mínimos Quadrados, temos:

$$\hat{\beta}_1 \sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} x_i y_i - \hat{\beta}_0 \sum_{i=1}^{n} x_i$$

$$= \sum_{i=1}^{n} x_i y_i - (\bar{Y} - \hat{\beta}_1 x) \sum_{i=1}^{n} x_i$$

$$= \sum_{i=1}^{n} x_i y_i - \bar{Y} \sum_{i=1}^{n} x_i - \bar{x} \hat{\beta}_1 \sum_{i=1}^{n} x_i$$

$$= \sum_{i=1}^{n} x_i y_i - n \bar{x} \bar{Y} + n \hat{\beta}_1 \bar{x}^2$$

$$\beta_1 \left(\sum_{i=1}^{n} x_i^2 - n \bar{x}^2 \right) = \sum_{i=1}^{n} x_i y_i - n \bar{x} \bar{Y}$$

$$\therefore \hat{\beta}_1 = \frac{\sum_{i=1}^{n} x_i y_i - n \bar{x} \bar{Y}}{\sum_{i=1}^{n} x_i^2 - n \bar{x}^2} = \frac{\sum_{i=1}^{n} (x_i - \bar{x}) Y_i}{\sum_{i=1}^{n} (x_i - \bar{x}) x_i} = \frac{\sum_{i=1}^{n} (x_i - x)(y_i - \bar{Y})}{\sum_{i=1}^{n} (x_i - \bar{x})^2}$$

Assim conseguimos obter e definir os estimadores $\hat{\beta}_0$ e $\hat{\beta}_1$, eles também são chamados de Estimadores de Mínimos Quadrados.

→ Suposições

Supomos que a distribuição dos erros siga uma distribuição normal e seu valor esperado seja o menor possível, no caso, 0. Supomos também que a variância seja constante e um método de checar se tal suposição é adequada, é pela técnica de estimação da variância residual.

→ Hipóteses

A hipótese nula define que não há correlação entre as variáveis X e Y:

$H_0: \beta_1 = 0$

A hipótese alternativa define que há correlação entre as nossas variáveis X e Y:

$H_1: \beta_1 \neq 0$

Caso a hipótese nula seja rejeitada, significa que a variável X interfere no resultado da variável Y. Caso ela não seja rejeitada, significa que existe a possibilidade de que a variável X não interfira na Y assim como há possibilidades de que ela interfira.