Modelo de Regressão Linear Simples

O modelo de regressão linear simples considera apenas uma variável explicativa e a função de regressão é linear. O modelo é definido por:

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i, \tag{1}$$

em que:

- Y_i é o valor da variável resposta para a i-ésima observação,
- β_0 é o intercepto e β_1 é o coeficiente angular, ambos são parâmetros desconhecidos,
- X_i é uma constante conhecida, o valor da variável explicativa para a i-ésima observação. X_i é uma variável fixa (ou sem erro ou determinística),
- ε_i são independentes e $N(0, \sigma^2)$.

Estimadores de $\hat{\beta}_0$, $\hat{\beta}_1$ e $\hat{\sigma}^2$

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n} X_i Y_i - \frac{(\sum_{i=1}^{n} X_i)(\sum_{i=1}^{n} Y_i)}{n}}{\sum_{i=1}^{n} X_i^2 - \frac{(\sum_{i=1}^{n} X_i)^2}{n}},$$

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X},$$

e

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2}{n-2} = \frac{\sum_{i=1}^{n} e_i^2}{n-2} = \frac{\text{SQRes}}{n-2} = \text{MSRes}.$$

Análise de Variância

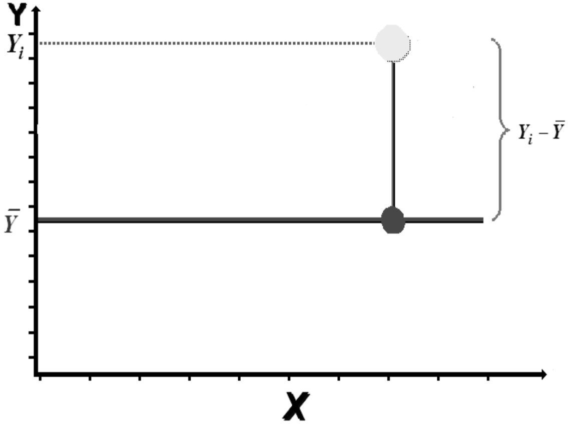

Figura : Desvio total

Análise de Variância

- Quando usamos a informação da variável explicativa X, a variação reflete a incerteza das observações Y_i relativa à reta de regressão ajustada:

$$Y_i - \hat{Y}$$

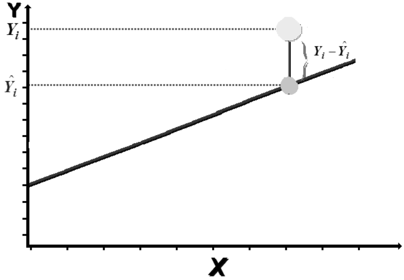

Figura : Desvios

- Ao passarmos do modelo reduzido (2) para o modelo de regressão linear (1), temos que:

$$SQT - SQRes,$$

essa diferença mede a redução na soma de quadrados devido ao uso do modelo de regressão linear e será definido por SQReg, soma de quadrados devido à regressão,

$$SQReg = SQT - SQRes = \sum_{i=1}^{n}(\hat{Y}_i - \bar{Y})^2 \tag{6}$$

Análise de Variância

- Da relação (6), podemos definir a soma de quadrados total por:

$$SQT = SQReg + SQRes \tag{7}$$

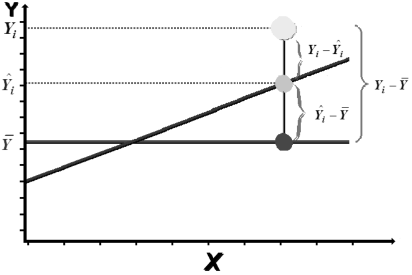

Figura : Decomposição da variação total

Análise de Variância

- O desvio total, $Y_i - \bar{Y}$, pode ser decomposto por:

$$(Y_i - \bar{Y}) = (\hat{Y}_i - \bar{Y}) + (Y_i - \hat{Y}_i) \tag{8}$$

- Os componentes são:
  1. Desvios dos valores estimados pela regressão em torno da média,
  2. Desvios em torno da reta de regressão.
- Essa decomposição tem a propriedade de que as somas desses desvios ao quadrado têm a mesma relação, ou seja,

$$\sum_{i=1}^{n}(Y_i - \bar{Y})^2 = \sum_{i=1}^{n}(\hat{Y}_i - \bar{Y})^2 + \sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2 \tag{9}$$

Análise de Variância

- A soma de quadrados dividida pelo respectivo grau de liberdade é denominada **Quadrado Médio**

- **Quadrado Médio da Regressão** - $MSReg = \frac{SQReg}{1}$

- **Quadrado Médio do Resíduo** - $MSRes = \frac{SQRes}{n-2}$

- Para fazer inferência baseada na metodologia de análise de variância, é necessário conhecer o valor esperado dos quadrados médios.

- $E(MSRes) = \sigma^2$

- $E(MSReg) = \sigma^2 + \beta_1^2 \sum_{i=1}^{n}(X_i - \bar{X})^2$

Análise de Variância

- Hipóteses:

$$H_0 : \beta_1 = 0$$
$$H_1 : \beta_1 \neq 0$$

ou,

$$H_0 : \text{ausência de regressão}$$
$$H_1 : \text{existe regressão}$$

Análise de Variância

- Estatística do teste

$$F = \frac{MSReg}{MSRes},$$

tem distribuição F com 1 grau de liberdade no numerador e (n-2) graus de liberdade no denominador.

- Se $F \leq F_{(1-\alpha,1,n-2)} \Rightarrow$ Há evidência para aceitar H_0
- Se $F > F_{(1-\alpha,1,n-2)} \Rightarrow$ Há evidência para rejeitar H_0

Coeficiente de Determinação

- SQT é a medida da variação das observações Y_i sem considerar o efeito da variável regressora X.
- SQRes é a medida da variação das observações Y_i quando o modelo de regressão utilizando a variável regressora X é considerado.
- Um medida do efeito de X em reduzir a variação de Y é

$$SQT - SQRes = SQReg$$

- Essa medida pode ser expressa em termos da proporção da variação total:

$$R^2 = \frac{SQReg}{SQT} = 1 - \frac{SQRes}{SQT}$$

Coeficiente de Determinação

- A medida R^2 é chamada de **Coeficiente de Determinação** e representa a proporção da variação total explicada pela relação X e Y (regressão).
- Desde de que $0 \leq SQRes \leq SQT$, segue que

$$0 \leq R^2 \leq 1$$

- Valores grandes de R^2 indicam que a variação total de Y é mais reduzida pela introdução da variável preditora X.
- Quando todas as observações estão na reta de regressão ajustada, SQRes=0 e $R^2 = 1$.
- Quando a reta de regressão ajustada é horizontal, então, $\hat{\beta}_1 = 0$ e SQRes=SQT. Logo, $R^2 = 0$.