

Regression Modeling

simple linear model

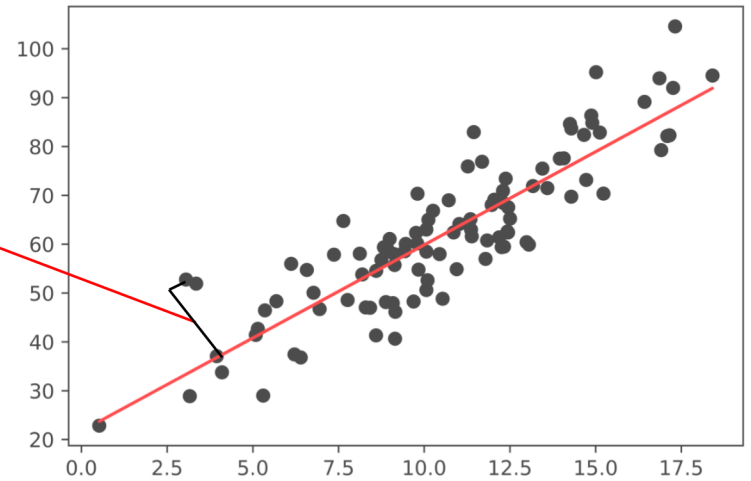
$$[Y_i = \beta_1 + \beta_2 X_{2i}] + \beta_3 X_{3i} + \dots + \beta_K X_{Ki} + u_i$$

Final outcome being modeled

slope intercept - value of "Y" without any additional input

β_2 - Multiplier for the effect that a change in X will have

u_i - Average distance between actual data points and the estimated regression line



Log transformations

- $\log(Y_i) = \beta_1 + \beta_2 \log(X_i) + u_i \rightarrow$

\rightarrow A 1% increase in X leads to a $\beta_2\%$ increase in Y

- $Y_i = \beta_1 + \beta_2 \log(X_i) + u_i \rightarrow$

\rightarrow A 1% increase in X leads to a $\beta_2/100$ increase in Y

- $\log(Y_i) = \beta_1 + \beta_2 X_i + u_i \rightarrow$

\rightarrow A one unit increase in X leads to a $\beta_2 * 100\%$ increase in Y

Useful Equations

$$\hat{\beta}_1 = \bar{Y} - \hat{\beta}_2 \bar{X} \quad \hat{\beta}_2 = \frac{\sum_{i=1}^N [(Y_i - \bar{Y})(X_i - \bar{X})]}{\sum_{i=1}^N [(X_i - \bar{X})^2]}$$

$$\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_i \quad \hat{u}_i = Y_i - \hat{Y}_i$$

$$TSS = ESS + RSS \quad R^2 = \frac{ESS}{TSS} = 1 - \frac{RSS}{TSS}$$

$$TSS = \sum_{i=1}^N [(Y_i - \bar{Y})^2] \quad ESS = \sum_{i=1}^N [(\hat{Y}_i - \bar{Y})^2] \quad RSS = \sum_{i=1}^N [\hat{u}_i^2]$$

$$t = \frac{\hat{\beta} - \beta^0}{SE\hat{\beta}} \quad \left[\hat{\beta} - SE(\beta) \cdot t\text{-crit}, \hat{\beta} + SE(\beta) \cdot t\text{-crit} \right]$$

Dummy Variables

Instances where "X" is either a 1 or 0

Ex: 0 if female, 1 if male

Often causes a change in slope/intercept

