

Regression and its Parameters

$$\hat{y} = \beta_0 + \beta_1 x$$

prediction

intercept

input variable

how much influence does x have on the prediction?

$$\beta_1 = \left(\frac{\sigma_y}{\sigma_x} \right) \cdot R$$

correlation between x and y

$$\beta_0 = \bar{y} - \beta_1 \bar{x}$$

\bar{x}, \bar{y} = sample mean of x and y

$$[\hat{y} = \beta_0 + \beta_1 x] \rightarrow (\bar{x}, \bar{y})$$

How do you know when to use a Linear Model?

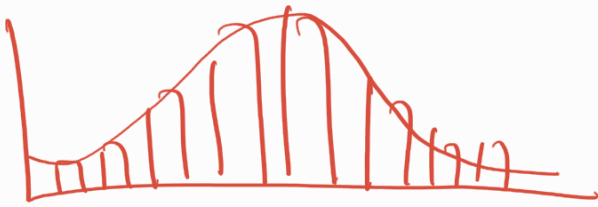
(1) LINEARITY

→ INPUT VARIABLES VS PREDICTED VARIABLE

(2) NEARLY NORMAL RESIDUALS

→ residual histogram should be nearly normally distributed centered around 0

→ non linearity in residuals can be an issue.



(3) CONSTANT VARIABILITY



homoscedasticity

How to use Categorical Variables in a Linear Model?

$$\hat{y} = \underbrace{\beta_0}_{\text{continuous}} + \underbrace{\beta_1}_{\text{continuous}} \underbrace{x}_{\text{continuous}}$$

$x = \text{Blood Group}$

is it 0^+ ?
is it 0^- ?
:
:

$O^+ = 0$
 $O^- = 1$
 $A^+ = 2$
:
:

0000001
0000010

Outliers

