# Simple Linear Regression

- Useful for Predicting quantitative Response
- Predicting $Y$ on the basis of a single predictor variable $X$
- Assuming that there is a linear relationship between $X$ and $Y$

$$Y \approx \beta_0 + \beta_1 X$$

- This can be read as regressing $Y$ on $X$
- or $Y$ onto $X$

  **Example :**
- TV ads $\rightarrow X$
- Sales $\rightarrow Y$
- Sales $\approx \beta_0 + \beta_1 TV$
  - $\beta_0$,$\beta_1$ two unknown constants
  - $\beta_0 \rightarrow$ Slop of $X$
  - $\beta_1 \rightarrow$ intercept of $Y$
  - They are called model **Coefficients** or **Parameters**

After Using Training data to estimate $\hat{\beta}_0, \hat{\beta}_1$

$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$$

# Estimating the Coefficients

- In practice $\beta_0, \beta_1$ are unknown, we usually useTraining Data to estimate them :$(x_1, y_1), \ldots, (x_n, y_n)$
- we try to estimate $\beta_0, \beta_1$ as close as possible to the data points so:
  - $yi \approx \hat{\beta}_0 + \hat{\beta}_1 x_i$
- To minimize as much as possible $\rightarrow$ we use Least squares criterion

Let $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_i \rightarrow e_i = y_i - \hat{y}_i$

- $y_i$ observed response
- $\hat{y}_i$ predicted response

  With that we have Residual Sum of Squares