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 $\approx \beta_0 + \beta_1 TV$
  - β_0,β_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 $\hat{\beta}_0, \hat{\beta}_1$

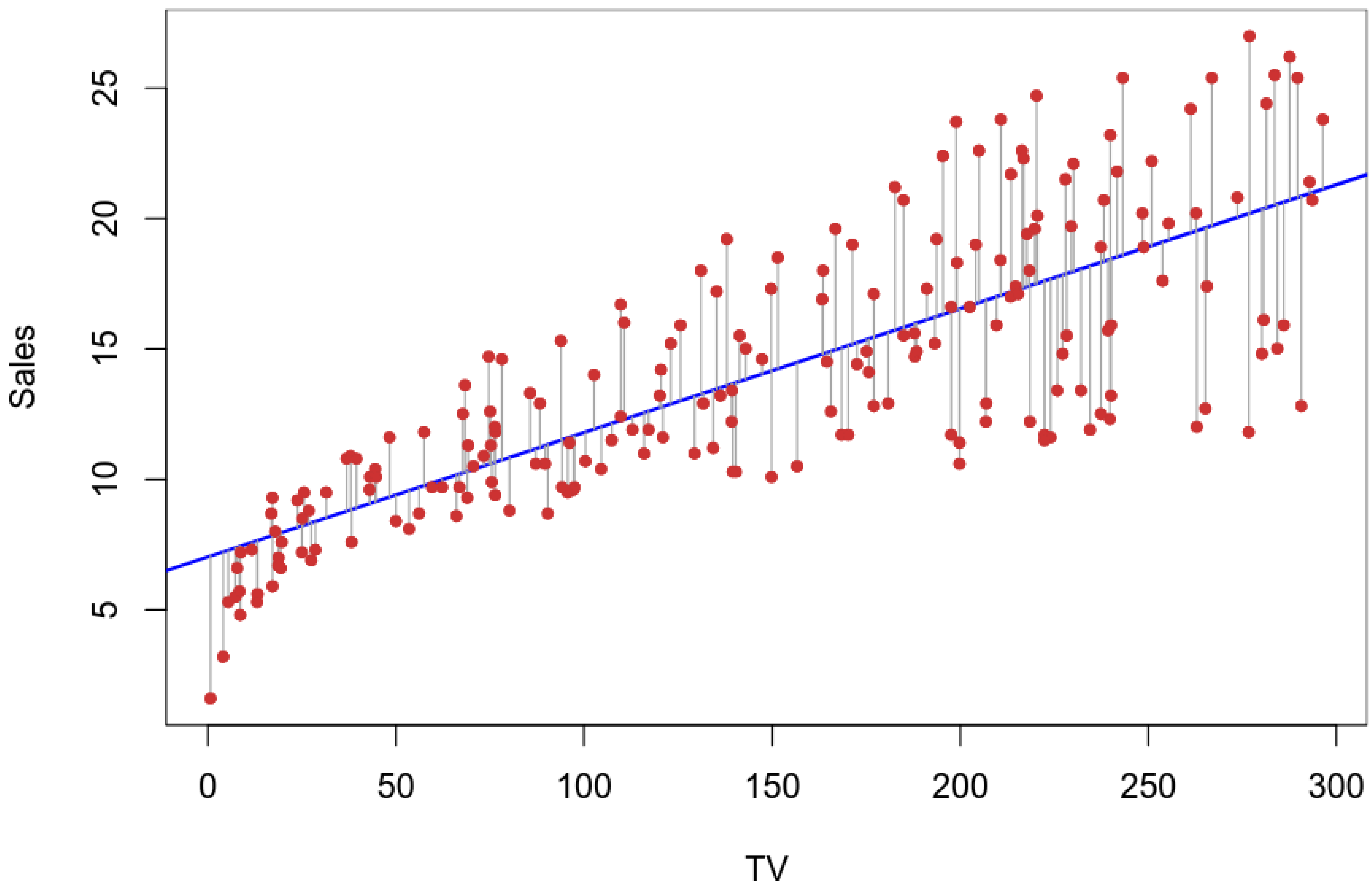
$$\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$$

Estimating the Coefficients

- In practice β_0, β_1 are unknown, we usually useTraining Data to estimate them :$(x_1, y_1), \ldots, (x_n, y_n)$
- we try to estimate β_0, β_1 as close as possible to the data points so:
  - $y_i \approx \hat{\beta}_0 + \hat{\beta}_1 x_i$
- 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 $\rightarrow$ true values in Training Data
- $\hat{y}_i$ predicted response $\rightarrow$ from the regression line
  With that we have Residual Sum of Squares (RSS)



- The Blue line fit is found by the lease squares
- Minimizing the residual sum of squares
    - Minimizing both β_0, β_1 Ordinary Least Squares
- Every grey line is the residual of $y_i - \hat{y}_i$

Assessing the Accuracy of the Coefficients β_0, β_1

1. Standard Error $SE(\hat{\beta}_1)$
2. Confidence Interval
3. Hypothesis Testing

- We assumed that the true relationship between X and Y is on form :

$$Y = f(X) + \varepsilon$$

- If f is approximated to be linear then we can write the relationship as:

$$Y = \beta_0 + \beta_1 x + \varepsilon$$

    - $\beta_1 \to$ The average increase in Y associated with one unite increase in X
- The population mean μ is usually unknown so the sample mean $\bar{\mu}$ will provide a good estimate to μ
- The sample mean $\bar{\mu}$ is unbiased $\to$ it averages out an estimation of huge biased estimations so that the sample mean $\bar{\mu}$ will be as close to the real population mean μ
  Using the same Logic we make multiple estimations for Coefficients $\hat{\beta}_0, \hat{\beta}_1$ and averaging it out will be spot on
- The question now is how a single estimation is far from the mean $\to \text{variance}$

$$\text{Standard error } \text{Var}(\hat{\mu}) = SE(\bar{\mu})^2 = \frac{\sigma^2}{n}$$

- σ is the standard deviation of each realization y_i

In the same tone we can see how close $\hat{\beta}_0, \hat{\beta}_1$ are to true values β_0, β_1

$$\mathrm{SE}(\hat{\beta}_o)^2 = \sigma^2 \left[\frac{1}{n} + \frac{\bar{x}^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \right]$$

$$\mathrm{SE}(\hat{\beta}_1)^2 = \frac{\sigma^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

Standard Error Derivation