Model : $Y = \beta_0 + \beta_1 X + \varepsilon_i$ $\qquad \varepsilon_i \sim N(0, \sigma^2)$

$$E[Y] = E[\beta_0 + \beta_1 X_i + \varepsilon_i]$$
$$= E[\beta_0 + \beta_1 X_i]$$
$$E[Y|X] = \beta_0 + \beta_1 X$$

$$\text{var}[Y] = \text{var}[\beta_0 + \beta_1 X + \varepsilon_i]$$
$$= \text{var}[\beta_0 + \beta_1 X] + \text{var}[\varepsilon_i] \quad \text{from independence.}$$
$$\text{Var}[Y|X] = \text{var}[\varepsilon_i] = \sigma^2$$

Sample of $\{(x_1, y_1) \ldots (x_n, y_n)\}$

Want to find best parameter to fit data?
$\hookrightarrow$ Least Squares of MLE

Model : $y_i = \beta_0 + \beta_1 X_i + \varepsilon_i$

What is β_0, β_1 ?

$$\hat{\beta} = \arg\min_{\beta_0, \beta_1} \sum_{i=1}^{n} \left(Y_i - (\beta_0 + \beta_1 X_i)\right)^2$$

If we have that $\beta_0 = 0$

$$\hat{\beta} = \arg\min_{\beta_0, \beta_1} \sum_{i=1}^{n} (Y_i - \beta X_i)^2 \quad \longrightarrow \quad \text{Take derivative}$$

$$\frac{d}{dx} \sum (Y_i - \beta X_i)^2 = \sum_{i=1}^{n} (Y - \beta X_i)(-2X_i)$$

$$\rightarrow \sum_{i=1}^{n} -2x_i (Y_i - \beta x_i) = 0$$

$$\rightarrow \sum 2\beta x_i^2 = 2 \sum x_i Y_i$$

$$\hat{\beta} = \frac{\sum x_i Y_i}{\sum x_i^2} \quad \longrightarrow \text{LS sol}^{n}$$

$\hat{\beta}$ is the slope estimate & a R.V

$$\hat{\beta} = \frac{\sum x_i y_i}{\sum x_i^2} \quad , \quad \hat{\beta} \sim N(\beta_{true}, ?)$$

if $\hat{\beta}$ unbiased $E[\hat{\beta}] = \beta_{true}$

$$E[\hat{\beta}] = \frac{1}{M} \sum_{i=1}^{M} \hat{\beta}_i$$

$$(?)^{1/2} = \left(\frac{1}{M-1} \sum_{i=1}^{M} \left(\hat{\beta}_i - E(\hat{\beta}) \right)^2 \right)^{1/2}$$

→ Estimate of the std error of $\hat{\beta}$

↓ Std deviation