How to estimate β_0, β_1?

$$\underset{\beta_0, \beta_1}{\text{minimize}} \ RSS = \sum_{i=1}^{n} \left[y_i - (\beta_0 + \beta_1 x_i) \right]^2$$

$$\begin{cases} \frac{\partial}{\partial \beta_0} RSS = 0 \\ \frac{\partial}{\partial \beta_1} RSS = 0 \end{cases} \text{ when } \begin{matrix} \beta_0 = \hat{\beta}_0 \\ \beta_1 = \hat{\beta}_1 \end{matrix} \Rightarrow \begin{cases} \sum_{i=1}^{n} 2(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)(-1) = 0 & (1a) \\ \sum_{i=1}^{n} 2(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)(-x_i) = 0 & (1b) \end{cases}$$

$$\sum_i y_i = \sum_i (\hat{\beta}_0 + \hat{\beta}_1 x_i) = n\hat{\beta}_0 + \hat{\beta}_1 \sum x_i \Rightarrow \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

$$\sum_i y_i x_i = \sum_i \left(\hat{\beta}_0 x_i + \hat{\beta}_1 x_i^2 \right) \Rightarrow \sum x_i y_i = (\bar{y} - \hat{\beta}_1 \bar{x}) \sum x_i + \hat{\beta}_1 \sum x_i^2$$

$$= n \bar{y} \bar{x} + \hat{\beta}_1 \left(\sum x_i^2 - n\bar{x}^2 \right)$$

Let $\bar{x} = \frac{1}{N} \sum_{i=1}^{N} x_i$, $\bar{y} = \frac{1}{N} \sum_{i=1}^{N} y_i$.

$$\Rightarrow \hat{\beta}_1 = \frac{\sum x_i y_i - n\bar{x}\bar{y}}{\sum x_i^2 - n\bar{x}^2}$$

$$\underbrace{\begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}}_{\underset{\sim}{y}} = \underbrace{\begin{bmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix}}_{X} \underbrace{\begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}}_{\underset{\sim}{\beta}} + \underbrace{\begin{bmatrix} e_1 \\ \vdots \\ e_n \end{bmatrix}}_{\underset{\sim}{e}} \qquad \underset{\underset{\sim}{\beta}}{\text{minimize}} \ \| \underset{\sim}{y} - X \underset{\sim}{\beta} \|^2$$

$$\underset{\sim}{y} = X \underset{\sim}{\beta} + \underset{\sim}{e}$$

$$\nabla_{\underset{\sim}{\beta}} RSS = \underbrace{2X^T (\underset{\sim}{y} - X\underset{\sim}{\beta}) \Big|_{\underset{\sim}{\beta} = \hat{\underset{\sim}{\beta}}}}_{\text{Matrix calculus}} = \underset{\sim}{0} \Rightarrow \underbrace{X^T \underset{\sim}{y} = X^T X \hat{\underset{\sim}{\beta}}}_{\text{Normal Eqn}} \Rightarrow \hat{\underset{\sim}{\beta}} = \underbrace{(X^T X)^{-1} X^T}_{\text{Pseudoinverse}} y$$

$$X^T X = \begin{bmatrix} n & \sum x_i \\ \sum x_i & \sum x_i^2 \end{bmatrix} \quad X^T y = \begin{bmatrix} \sum y_i \\ \sum x_i y_i \end{bmatrix} \quad (X^T X)^{-1} = \frac{1}{n \sum x_i^2 - (\sum x_i)^2} \begin{bmatrix} \sum x_i^2 & -\sum x_i \\ -\sum x_i & n \end{bmatrix}^T$$

$$\hat{\beta} = \frac{1}{n\sum x_i^2 - (\sum x_i)^2} \begin{bmatrix} \sum x_i^2 & -\sum x_i \\ -\sum x_i & n \end{bmatrix} \begin{bmatrix} \sum y_i \\ \sum x_i y_i \end{bmatrix}$$

$$= \frac{1}{\sum x_i^2 - n\bar{x}^2} \begin{bmatrix} \frac{1}{n}\sum x_i^2 & -\bar{x} \\ -\bar{x} & 1 \end{bmatrix} \begin{bmatrix} \sum y_i \\ \sum x_i y_i \end{bmatrix}$$

$$= \frac{1}{\sum x_i^2 - n\bar{x}^2} \begin{bmatrix} \sum x_i^2 \bar{y} - \bar{x} \sum x_i y_i \\ -n\bar{x}\bar{y} + \sum x_i y_i \end{bmatrix} = \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{bmatrix}$$

Try to prove yourself:

$$\sum x_i^2 - n\bar{x}^2 = \sum (x_i - \bar{x})^2$$

$$\sum x_i y_i - n\bar{x}\bar{y} = \sum (x_i - \bar{x})(y_i - \bar{y})$$

* The matrix approach is more elegant, easier to implement and to interpret.

$$\mathrm{VarCov}(\hat{\underset{\sim}{\beta}}) = \mathrm{VarCov}\left[(X^TX)^{-1}X^T\underset{\sim}{y}\right]$$

$$= (X^TX)^{-1}X^T \underbrace{\mathrm{VarCov}(\underset{\sim}{y})}_{\sigma^2 \mathbb{I}} X(X^TX)^{-1} = \sigma^2 (X^TX)^{-1}$$

$$= \sigma^2 \begin{bmatrix} \frac{1}{n}\sum x_i^2 & -\bar{x} \\ -\bar{x} & 1 \end{bmatrix} \Big/ \sum (x_i - \bar{x})^2 = \begin{bmatrix} \mathrm{Var}(\hat{\beta}_0) & \mathrm{Cov}(\hat{\beta}_1, \hat{\beta}_0) \\ \mathrm{Cov}(\hat{\beta}_0, \hat{\beta}_1) & \mathrm{Var}(\hat{\beta}_1) \end{bmatrix}$$

$$\mathrm{Var}(\hat{\beta}_0) = \sigma^2 \cdot \frac{1}{n}\sum x_i^2 \cdot \frac{1}{\sum (x_i - \bar{x})^2} \quad \checkmark$$

$$\mathrm{Var}(\hat{\beta}_1) = \sigma^2 \Big/ \left(\sum x_i - \bar{x}\right)^2 \quad \checkmark$$

$$\underset{\sim}{X} = (X_0, \dots, X_n) \quad \underset{\sim}{W} = \underset{\sim}{X} - \mathbb{E}[\underset{\sim}{X}]$$

$$\mathrm{VarCov}(\underset{\sim}{X}) \overset{\text{def}}{=} \mathbb{E}[\underset{\sim}{W}\underset{\sim}{W}^T]$$

$$= \begin{bmatrix} \mathbb{E}[W_1^2] & & \mathbb{E}[W_n W_1] \\ & \ddots & \\ \mathbb{E}[W_1 W_n] & & \mathbb{E}[W_n^2] \end{bmatrix} = \begin{bmatrix} \mathrm{Var}(X_1) & & \mathrm{Cov}(X_1, X_n) \\ & \ddots & \\ \mathrm{Cov}(X_1, X_n) & & \mathrm{Var}(X_n) \end{bmatrix}$$