Lecture 4 Sept 29

Pop 1 $N(\mu_1, \sigma^2)$ $Y^1_1, \ldots, Y^1_{n_1}$ iid $N(\mu_1, \sigma^2)$

Pop 2 $N(\mu_2, \sigma^2)$ $Y^2_1 \ldots Y^2_{n_2}$ iid $N(\mu_2, \sigma^2)$

$$\delta = \mu_2 - \mu_1$$

$$T = \frac{(\bar{Y}_1 - \bar{Y}_2)}{\sqrt{S_p^2 \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$ Test Statistic

$$T \sim t(df = n_1 + n_2 - 2)$$ Under H_0

(Con't. page 4)

Recall: $\bar{Y}_1 - \bar{Y}_2 \sim N\left(\delta, \sigma^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right)\right)$

$\forall \delta$

$$\Rightarrow \quad \text{(A)} \quad \frac{(\bar{Y}_1 - \bar{Y}_2) - \delta}{\sqrt{\sigma^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} \sim N(0,1)$$

$$\text{(B)} \quad \frac{(n_1+n_2-2)\, S_p^2}{\sigma^2} \sim \chi^2(df = n_1+n_2-2)$$

$$\text{(C)} \quad S_p^2 \equiv g(S_1^2, S_2^2)$$

$$(\bar{Y}_1, \bar{Y}_2) \perp S_p^2$$

$$\Rightarrow \quad \text{(D)} \quad \frac{\dfrac{(\bar{Y}_1 - \bar{Y}_2) - \delta}{\sqrt{\cancel{\sigma^2}\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}}{\sqrt{\dfrac{\cancel{(n_1+n_2-2)} S_p^2}{\cancel{\sigma^2}} \Big/ \cancel{(n_1+n_2-2)}}} \sim t(n_1+n_2-2)$$

$$\frac{(\bar{Y}_1 - \bar{Y}_2) - \delta}{\sqrt{S_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \sim t(n_1 + n_2 - 2)$$

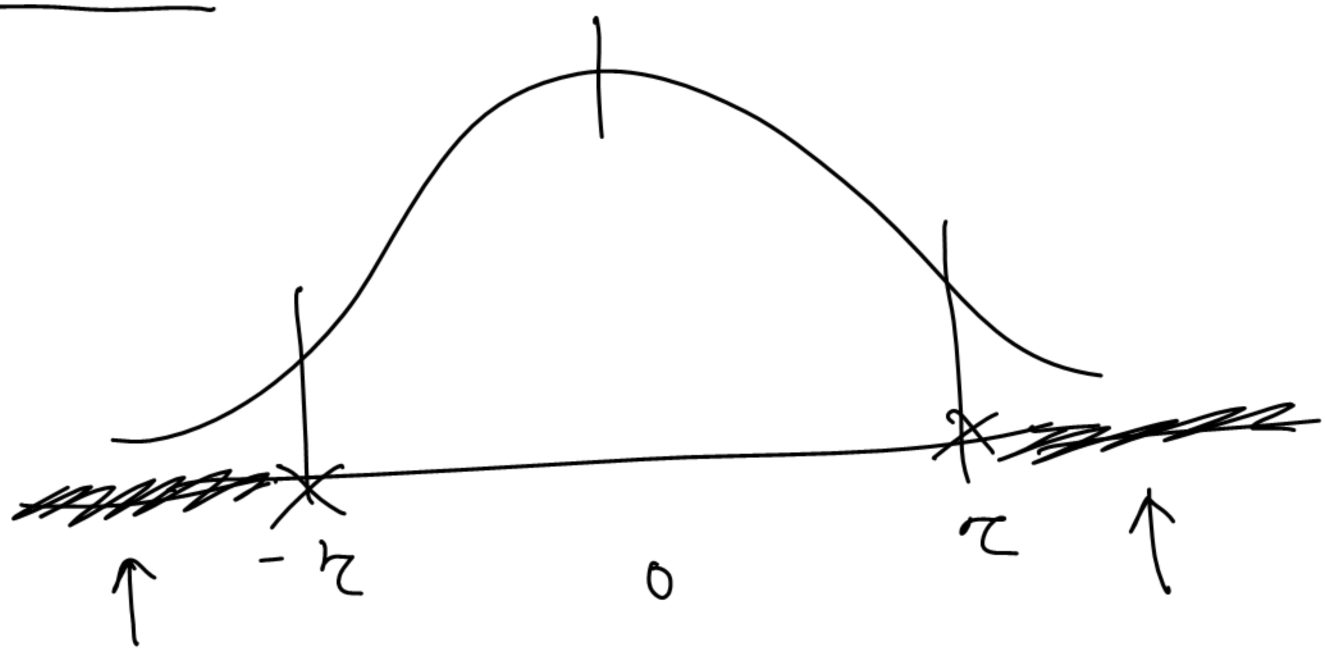
Under H_0: $\delta = 0$

So, when H_0 is true:

$$\frac{\bar{Y}_1 - \bar{Y}_2}{\sqrt{S_p^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \sim t(n_1 + n_2 - 2)$$

CON'T.

$t\ (df = n_1 + n_2 - 2)$

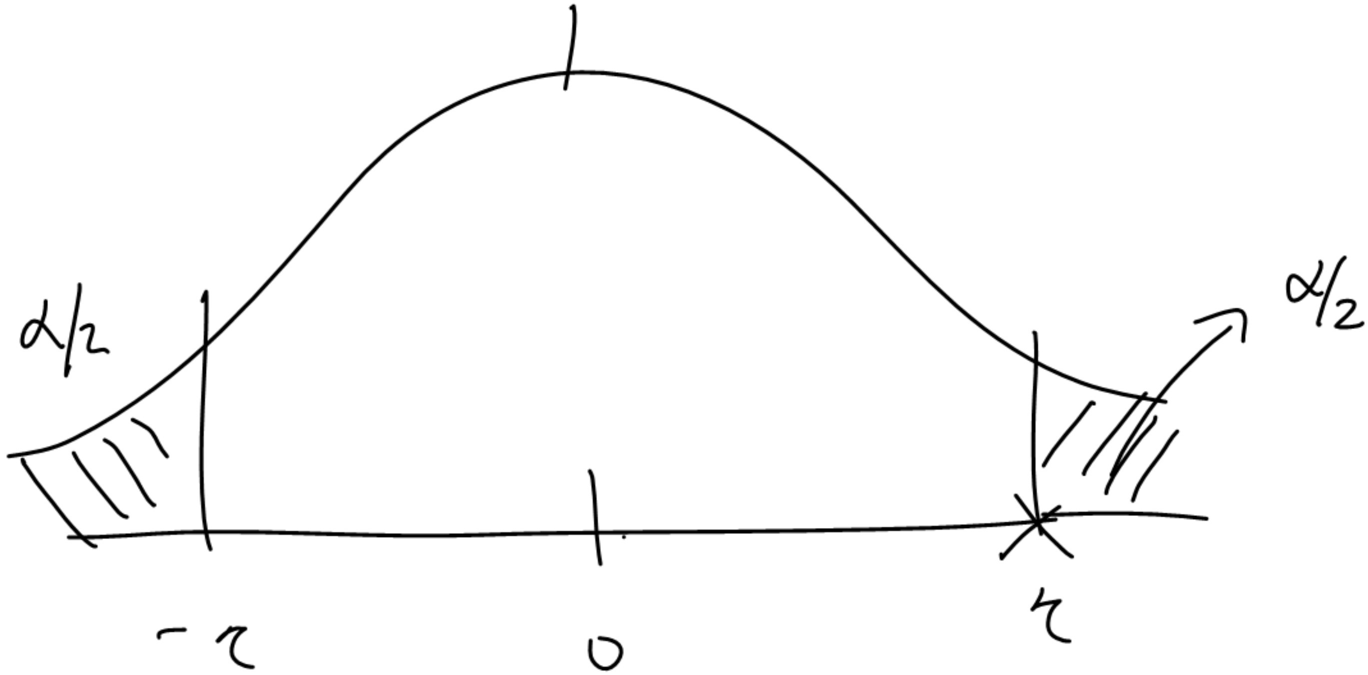


Find τ st we control the probability of Type I error to be a small quantity say α

$$P(\underbrace{\text{Reject } H_0 \mid H_0 \text{ true}}_{\text{Type I error}}) = \alpha$$

$$P(\,|T| > t \mid \delta = 0\,) = \alpha$$

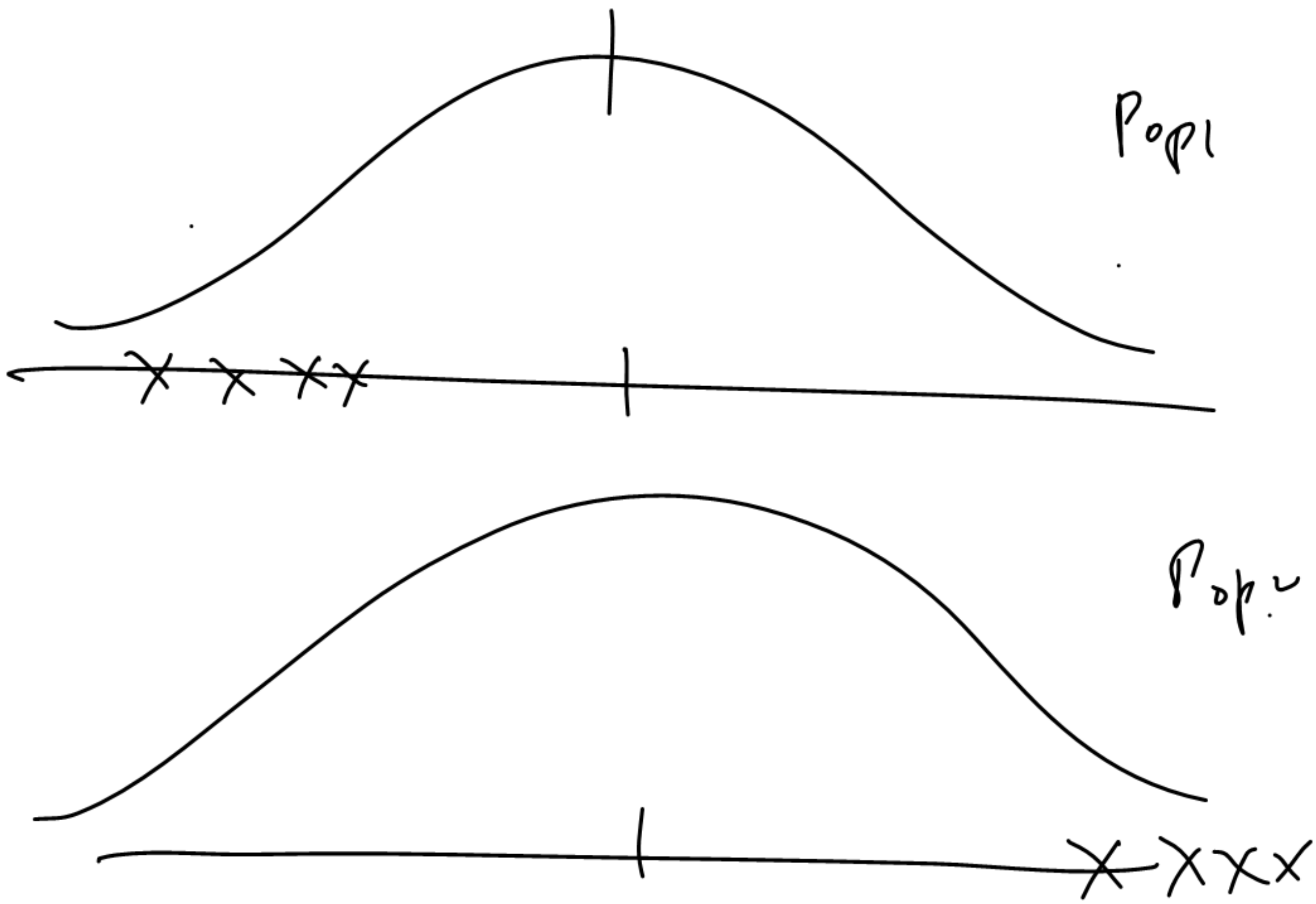
Since $T \overset{H_0 : \delta = 0}{\sim} t(n_1 + n_2 - 2)$



$\alpha = .05$

$\tau = (1 - \alpha/2) \times 100^{\text{th}}$ percentile of $t(n_1 + n_2 - 2)$

$= 97.5^{\text{th}}$ percentile of $t(n_1 + n_2 - 2)$



Here $\mu_1 = \mu_2$

So far • Two independent sample

- t-stat, behavior under H_0
- linear model framework
- likelihood function

• Paired t-test

Pair i $(Y_i^1, Y_i^2) \longrightarrow Z_i = Y_i^1 - Y_i^2$

$i = 1, \dots, n \longrightarrow Z_1, \dots, Z_n$

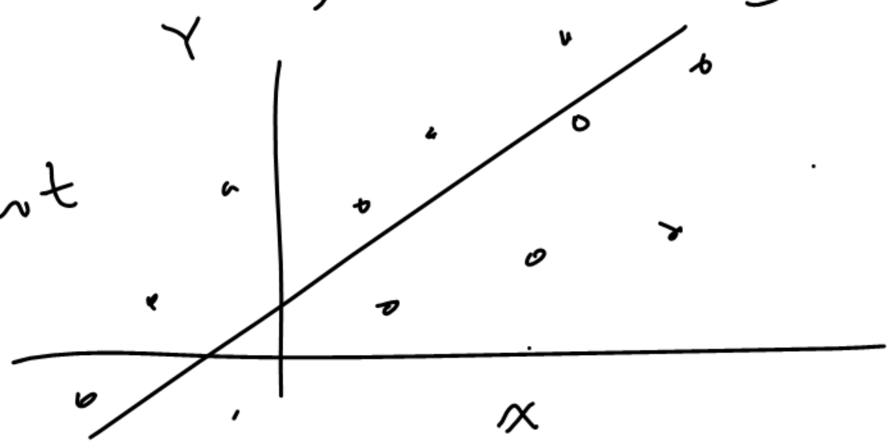
$\boxed{\text{Linear Regression}}$

Y_i = dependent / outcome variable

x_i = indep / predictor

DATA: $\{(x_i, Y_i),\ i = 1, \dots, n\}$

Scatterplot



Trend ?

Variation around the trend ?

Linear Regression Model

$$Y_i \mid x_i \sim N(\mu(x_i), \sigma^2)$$

(☆)

$$\mu(x_i) = \beta_0 + \beta_1 x_i$$

(☆☆)

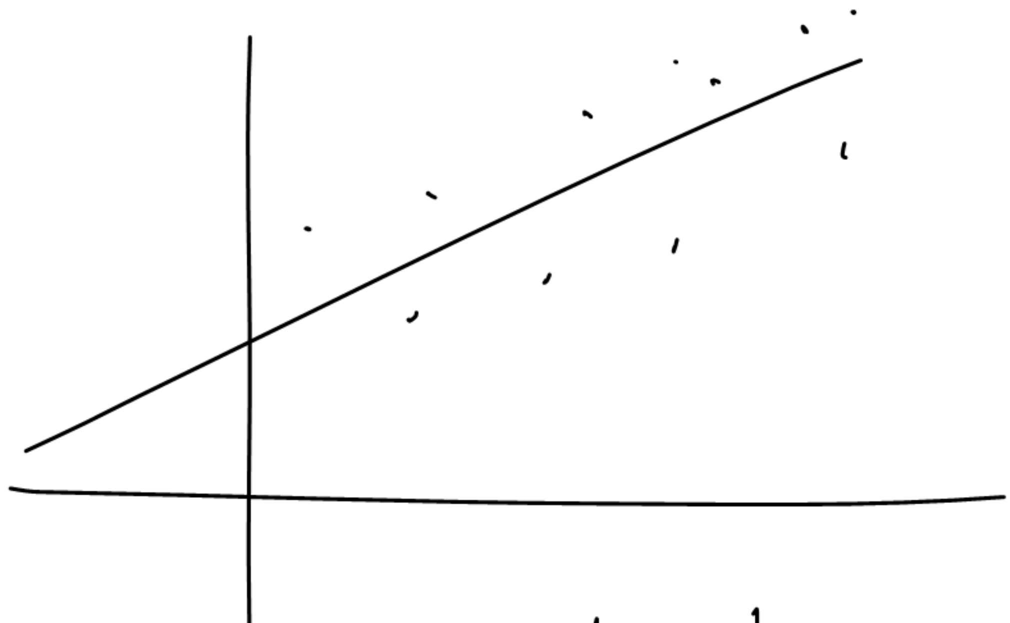
$$Y_i = \mu(x_i) + \varepsilon_i$$

constant over i

where $\varepsilon_i \sim N(0, \sigma^2)$

$$\underbrace{\mu(x_i)}_{\text{Trend}} = \beta_0 + \beta_1 x_i$$

σ^2 = variation around the trend



Estimate β_0, β_1

intercept — β_0; slope — β_1

σ^2 — variation around the trend

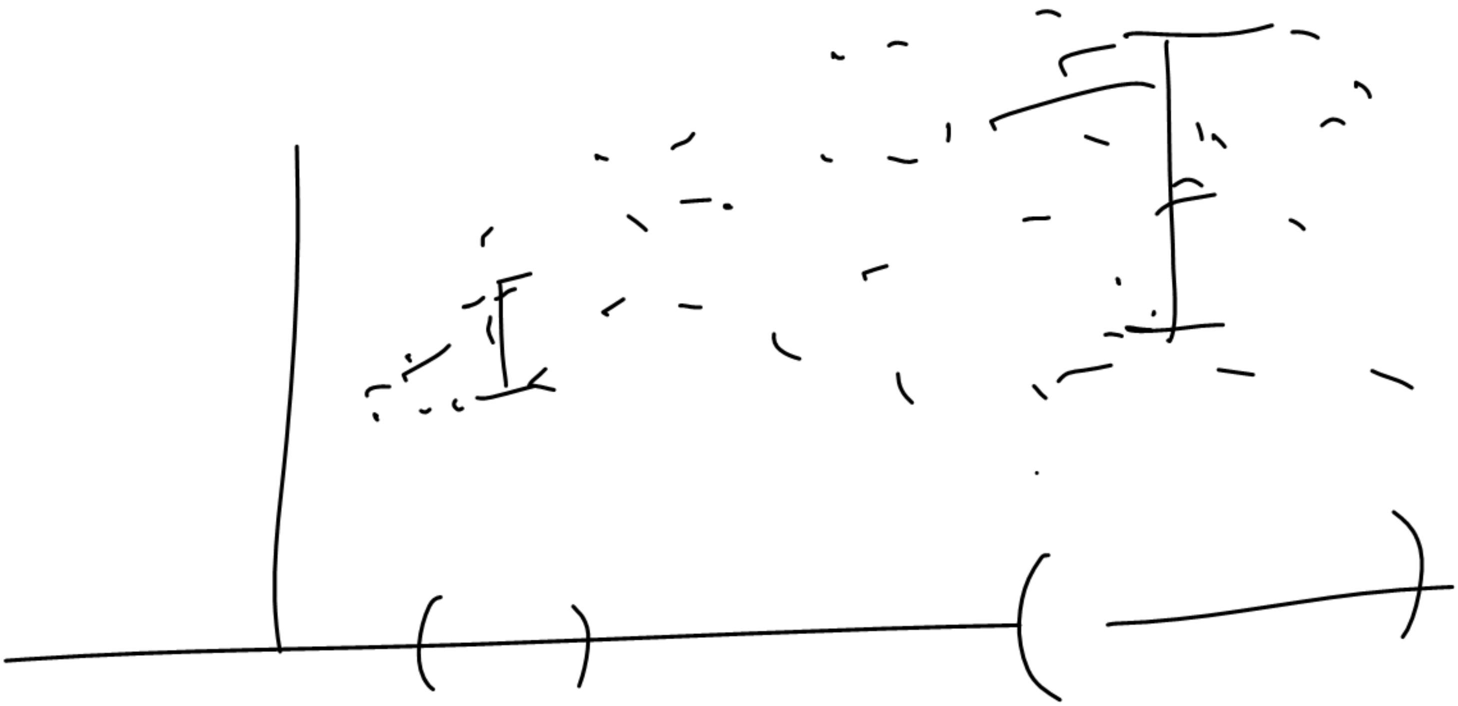
Note: ☆ $Var(Y_i \mid x_i) = \sigma^2$

☆☆ $Var(Y_i \mid x_i) = Var(\mu(x_i) + \varepsilon_i)$

$= Var(\varepsilon_i)$

$= \sigma^2$

variation around the trend is fixed over all x's.

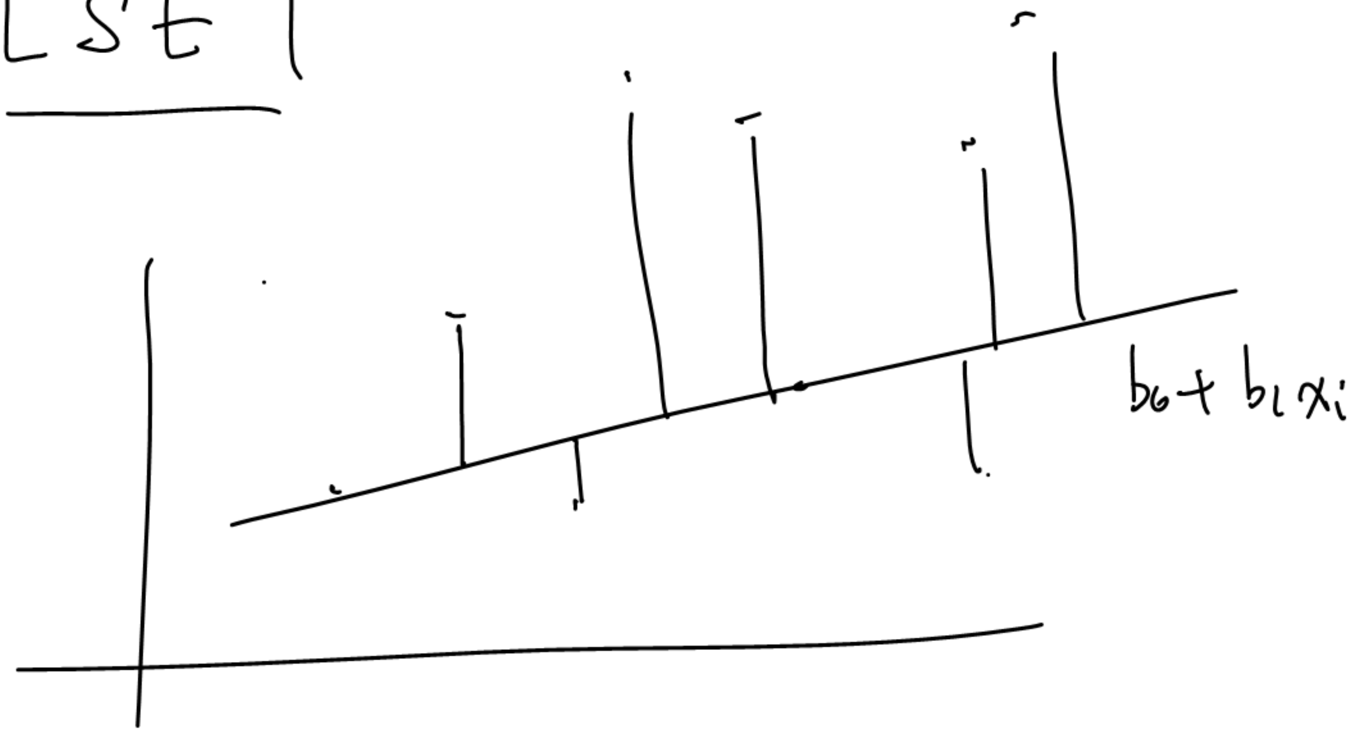


Estimate β_0, β_1

(1) Least Squares Estimation

(2) Maximum Likelihood

LSE



Let b_0 and b_1 be candidate estimates of β_0 and β_1.

The corresponding candidate trend is $b_0 + b_1 x_i$. Thus

$$C(b_0, b_1) = \sum_{i=1}^{n} \left(Y_i - (b_0 + b_1 x_i) \right)^2$$

is the corresponding squared error of the candidate trend.

The least squares estimator of β_0 & β_1, denoted by $\tilde{\beta}_0$ and $\tilde{\beta}_1$ satisfy:

$$(\tilde{\beta}_0, \tilde{\beta}_1) = \underset{(b_0, b_1) \in \mathbb{R}^2}{\arg\min} \; C(b_0, b_1)$$

To derive $\tilde{\beta}_0$ & $\tilde{\beta}_1$:

$$\frac{\partial C(b_0, b_1)}{\partial b_0} = -2 \sum_{i=1}^{n} \left(Y_i - (b_0 + b_1 x_i)\right)$$

$$\frac{\partial C(b_0, b_1)}{\partial b_1} = -2 \sum_{i=1}^{n} x_i \left(Y_i - (b_0 + b_1 x_i)\right)$$

$$\left.\frac{\partial \underline{C}(b_0, b_1)}{\partial \underline{b}}\right|_{(\tilde{\beta}_0, \tilde{\beta}_1)} = \underline{0}$$

$$\Rightarrow \cancel{-2} \sum \left(Y_i - (\tilde{\beta}_0 + \tilde{\beta}_1 x_i)\right) = 0$$

$$\cancel{-2} \sum \left(Y_i - (\tilde{\beta}_0 + \tilde{\beta}_1 x_i)\right) x_i = 0$$

SOLVE:

$$\Longrightarrow \tilde{\beta}_0 \quad , \quad \tilde{\beta}_1$$

In matrix notation: $Y_i = \mu(x_i) + \varepsilon_i$

$= (\beta_0 + \beta_1 x_i) + \varepsilon_i$

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ & \\ 1 & x_n \end{bmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix} + \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

$$\underline{Y} = X \underline{\beta} + \underline{\varepsilon}$$

Candidate estimate $\underline{b}$

Squared $C(\underline{b}) = \| \underline{Y} - X\underline{b} \|^2$

$= (\underline{Y} - X\underline{b})^T (\underline{Y} - X\underline{b})$

$\hat{\beta} = (X'X)^{-1}X'\underline{Y}$ <u>derive</u>!