Lecture 5 Oct 09

"Random" Data set

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

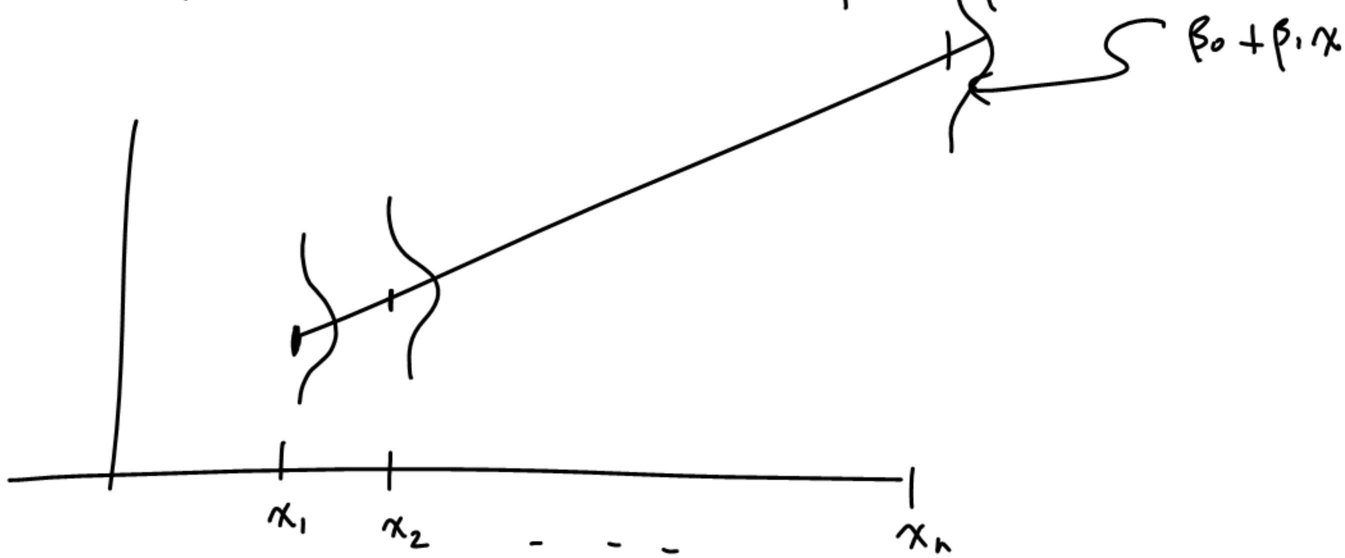
"Observed" Data set

$$\{ (x_i, y_i),\ i = 1 \dots, n \}$$

Model A

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

where $\mu(x_i) = E(Y_i \mid x_i) = \beta_0 + \beta_1 x_i$



Model B

$$Y_i = \mu(x_i) + \varepsilon_i \quad \text{where} \quad \varepsilon_i \overset{\text{indep}}{\sim} N(0, \sigma^2)$$

$$\text{and} \quad \mu(x_i) = \beta_0 + \beta_1 x_i$$

Rem. Model A and Model B are equivalent!

Model B

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

$$\underline{Y} = \underbrace{X}_{\text{design matrix}} \underbrace{\underline{\beta}}_{\text{parameter vector}} + \underline{\varepsilon}$$

<u>Goals</u>: (a) Estimate $\underline{\beta}$

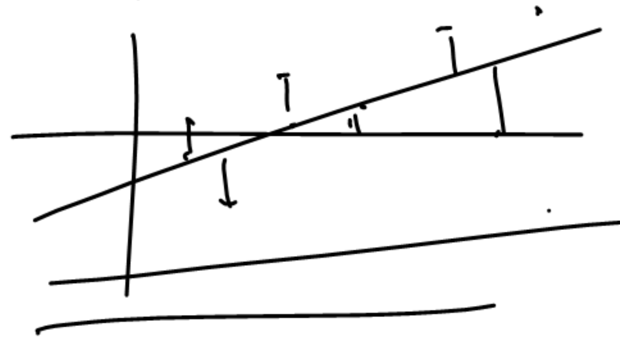
(b) Inference on $\underline{\beta}$ e.g. (i) $H_0: \beta_1 = 0$ vs $H_1: \beta_1 \neq 0$

(ii) Confidence interval for $E(Y|x^*) = \mu(x^*)$

(iii) Prediction interval for Y when $x = x^*$

To estimate $\underline{\beta}$ — Least Squares Estimation / Maximum Likelihood

Least Squares Estimation Let b_0 & b_1 be candidate estimators. The squared error for (b_0, b_1) is:



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

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

$$\frac{\partial C(b_0,b_1)}{\partial b_1} = -2\sum\left[Y_i-(b_0+b_1 x_i)\right]x_i \quad \text{——— } (**)$$

The LSE of β, denoted by $\hat{\beta}=\begin{pmatrix}\hat{\beta}_0\\ \hat{\beta}_1\end{pmatrix}$, satisfies:

$$\left.\begin{matrix}\dfrac{\partial C}{\partial b_0}\\[2ex] \dfrac{\partial C}{\partial b_1}\end{matrix}\right|_{\hat{\beta}=\begin{pmatrix}\hat{\beta}_0\\ \hat{\beta}_1\end{pmatrix}} = \begin{pmatrix}0\\0\end{pmatrix}$$

$$\sum_{i=1}^{n} Y_i = \sum_{i=1}^{n} (\hat{\beta}_0 + \hat{\beta}_1 x_i) \quad \text{from } (*) \text{ and}$$

$$\Rightarrow \quad \sum x_i Y_i = \sum_{i=1}^{n} (\hat{\beta}_0 + \hat{\beta}_1 x_i) x_i \quad \text{from } (**)$$

$$\Rightarrow \quad \sum Y_i = n\hat{\beta}_0 + \hat{\beta}_1 \sum x_i$$

$$\sum x_i Y_i = \hat{\beta}_0 \sum x_i + \hat{\beta}_1 \sum x_i^2$$

Proceed by elimination

$$\Rightarrow \quad \sum x_i \sum Y_i = n \sum x_i \hat{\beta}_0 + \hat{\beta}_1 \left(\sum x_i\right)^2$$

$$-n \sum x_i Y_i = -n \sum x_i \hat{\beta}_0 - n \sum x_i^2 \hat{\beta}_1$$

$$\Rightarrow \quad \left(\sum x_i \sum Y_i\right) - n \sum x_i Y_i = \hat{\beta}_1 \left(\left(\sum x_i\right)^2 - n \sum x_i^2\right)$$

$$\Rightarrow \quad \hat{\beta}_1 = \frac{\sum x_i \sum Y_i - n \sum x_i Y_i}{\left(\sum x_i\right)^2 - n \sum x_i^2}$$

$$= \frac{n \sum x_i Y_i - \sum x_i \sum Y_i}{n \sum x_i^2 - \left(\sum x_i\right)^2}$$

<u>Note</u> :

- $$n \sum x_i^2 - (\sum x_i)^2 = n\Big[\sum x_i^2 - \frac{1}{n}\underbrace{(\sum x_i)^2}\Big]$$

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

$$= n\Big[\sum x_i^2 - \frac{1}{n}\cdot n^2\bar{x}^2\Big]$$

$$= n\Big[\underbrace{\sum x_i^2 - n\bar{x}^2}\Big]$$

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

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

$$\vdots$$

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

<u>Claim</u>: $n\left(\sum (x_i - \bar{x})(y_i - \bar{y})\right) = n\left(\sum x_i y_i - n\bar{x}\bar{y}\right)$

LHS: $$n\left(\sum (x_i - \bar{x})(y_i - \bar{y})\right) = n\Big[\sum x_i y_i - \underbrace{x_i\bar{y}} - \underbrace{\bar{x}y_i} + \bar{x}\bar{y}\Big]$$

$$= n\Big[\sum x_i y_i - n\bar{x}\bar{y} \cancel{- n\bar{x}\bar{y}} \cancel{+ n\bar{x}\bar{y}}\Big]$$

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

$$\hat{\beta}_1 = \frac{\sum (x_i - \bar{x})(Y_i - \bar{Y})}{\sum (x_i - \bar{x})^2} \qquad \begin{array}{l} \leftarrow \operatorname{cov}(X, Y) \\[2em] \leftarrow \operatorname{var}(X) \end{array}$$

Another form for $\hat{\beta}_1$: the numerator can be expressed as:

$$\begin{aligned}
\sum (x_i - \bar{x})(Y_i - \bar{Y}) &= \sum (x_i - \bar{x}) Y_i - \sum (x_i - \bar{x}) \bar{Y} \\
&= \sum (x_i - \bar{x}) Y_i - \bar{Y} \left(\cancelto{0}{\sum (x_i - \bar{x})} \right) \\
&= \sum (x_i - \bar{x}) Y_i
\end{aligned}$$

$$\therefore \quad \hat{\beta}_1 = \frac{\sum (x_i - \bar{x}) Y_i}{\underbrace{\sum (x_j - \bar{x})^2}_{A}}$$

$$= \sum_{i=1}^{n} \frac{(x_i - \bar{x})}{A} Y_i$$

$$= \sum_{i=1}^{n} w_i Y_i \quad \text{linear combination of } \{Y_i\}$$

//

where the weights $\{w_i\}$:

$$w_i = \frac{x_i - \bar{x}}{\sum (x_j - \bar{x})^2} \quad //$$

In Summary, the LSE of β_1 is:

$$\hat{\beta}_1 = \frac{\sum (x_i - \bar{x})(Y_i - \bar{Y})}{\sum (x_i - \bar{x})^2}$$

$$= \sum w_i Y_i, \quad w_i = \frac{x_i - \bar{x}}{\sum_j (x_j - \bar{x})^2}$$

$\boxed{\text{Estimator}}$ $\quad$ $\boxed{\text{Estimate}}$

$\{(x_i, Y_i)\}$ $\quad$ $\{(x_i, y_i)\}$

The LS estimate of β_1 is:

$$\hat{\beta}_{1,obs} = \sum w_i \cdot y_i$$

The LSE of β_0 is:

$$\hat{\beta}_0 = \frac{\Sigma Y_i - \hat{\beta}_1 \Sigma x_i}{n}$$

$$= \bar{Y} - \hat{\beta}_1 \bar{x}$$

The LS estimate of β_0 is:

$$\hat{\beta}_{0,obs} = \bar{y} - \hat{\beta}_1 \bar{x}$$

The estimator for the best line:

$$\hat{\beta}_0 + \hat{\beta}_1 \cdot x_i$$

derived!

An estimator for σ^2 to be:

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{n} \left(Y_i - (\hat{\beta}_0 + \hat{\beta}_1 x_i)\right)^2}{\underline{\underline{n}} - (2)}$$