L1. Simple Linear Regression (SLR)

- Overview of Statistical Regression Models and Machine Learning
- SLR : model set up
- SLR : least square estimate of β_0 and β_1
- SLR : mean and variance of β_0 and β_1

I. Overview of Statistical Regression Models and supervised Machine Learning models.

Regression Models $\xrightarrow[\text{More features}]{\text{More data}}$ ML

↓ (Regression Models)

More about inferences

- relationship between variables
- significant feature
- inferences from the data.

↓ (ML)

More about prediction

- compare multiple approaches
- use train and test
- higher prediction accuracy

Improve performance and deal with complexity

Foundation and Interpretation

Strategy of regression Models

Statement of problem under consideration

↓

Data collection on relevant variables

↓

EPA : Initial check

↓

Develop one or more regression models

↓

Model Validation and criticism

↓

Identify the "best" model

↓

Model Interpretation and prediction

II. SLR : Model set-up

1. Classic Simple Linear Regression Model

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \quad , \quad i = 1, \cdots, n$$

- y_i : i^{th} value of the response variable y
- x_i : i^{th} value of the predictor / independent variable X
- n : total number of observations
- β_0 : intercept parameter (when $x = 0$, $y = \beta_0$)
- β_1 : slope parameter (when $x \uparrow 1$, $y \uparrow \beta_1$)
- ε_i : random error

NOTE

1. Classic: related to model assumptions are "standard"
2. Simple: only one predictor X
3. Linear: y is linear to parameters β_0 and β_1

EXAMPLE: $y_i = \beta_0 + \beta_1^2 / \beta_0 \, x_i + \varepsilon_i$ — non linear

$y_i = \beta_0 + \beta_1 x_i^2 + \varepsilon_i$ — linear.

2. Model Assumptions

- ε_i's are the RANDOM error terms that satisfy

  (i) $E(\varepsilon_i) = 0$

  (ii) $Var(\varepsilon_i) = \sigma^2$ — constant

  (iii) $Cov(\varepsilon_i, \varepsilon_j) = 0$ — no correlations between errors

  error of one obs. can't predictive the error of another obs.

  (iv) CLASSIC assumption:

  $\varepsilon_i \overset{i.i.d.}{\sim} N(0, \sigma^2)$ - Not required for LSE

  - needed for model Inference.

- X is a FIXED effect to Y

  (i) X_i's are treated as individual constants.

  (ii) We don't need to worry about X's distribution.

  — There is a mixed effect model when you need to include random effects. Not discussed in this course.

3. Regression Function.

The regression function is the mean value $E(y_i)$ of the response variable y given x_i, <u>the goal of SLR is to estimate $E(y_i)$</u>

Recall : $\underset{\text{random}}{y_i} = \underbrace{\beta_0 + \beta_1 x_i}_{\text{fixed}} + \underbrace{\varepsilon_i}_{\text{random}}$

• Regression Function: $E(y_i) = E(\beta_0 + \beta_1 x_i) + E(\varepsilon_i)$

$$= (\beta_0 + \beta_1 x_i) + 0$$

$$\boxed{E(y_i) = \beta_0 + \beta_1 x_i}$$

• $Var(y_i) = Var(\beta_0 + \beta_1 x_i + \varepsilon_i) = Var(\varepsilon_i) = \sigma^2$

• $cov(y_i, y_j) = cov(\beta_0 + \beta_1 x_i + \varepsilon_i, \beta_0 + \beta_1 x_j + \varepsilon_j)$

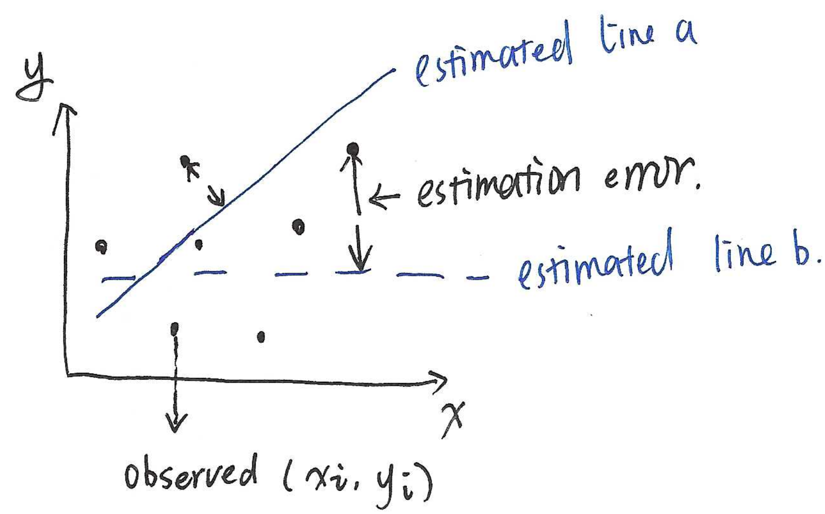
$$= cov(\varepsilon_i, \varepsilon_j)$$

$$= 0$$

III. SLR : least-square estimate of β_0 and β_1

Goal: we want to estimate $E(y_i) = \beta_0 + \beta_1 x_i$, $\Longleftrightarrow$ estimate β_0 and β_1

estimating principle: minimize the error between the observed y_i and estimated $\hat{y}_i = \hat{E}(y_i)$



1. Construct the deviation between the observed response y_i and the to-be-estimated linear mean component:

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

2. Construct a loss function to evaluate the "overall" error; here the classic choice is the $\boxed{\text{sum of squares}}$

$$S(\beta_0, \beta_1) = \sum_{i=1}^{n} \varepsilon_i^2 = \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i)^2$$

3. Choose the values $(\hat{\beta}_0, \hat{\beta}_1)$ that minimize the loss function:

$$(\hat{\beta}_0, \hat{\beta}_1) = \operatorname{argmin} S(\beta_0, \beta_1)$$

Solution:

$$\begin{cases} \dfrac{\partial S}{\partial \beta_0} = 0 \\ \dfrac{\partial S}{\partial \beta_1} = 0 \end{cases} \Rightarrow \begin{cases} -2\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i) = 0 \\ -2\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i)\, x_i = 0 \end{cases}$$

$$\Rightarrow \begin{cases} \sum y_i - n\beta_0 - \beta_1 \sum x_i = 0 \\ \sum y_i x_i - \beta_0 \sum x_i - \beta_1 \sum x_i^2 = 0 \end{cases}$$

$$\Rightarrow \begin{cases} n\bar{y} - n\beta_0 - n\beta_1 \bar{x} = 0 & (1) \\ \sum x_i y_i - n\beta_0 \bar{x} - \beta_1 \sum x_i^2 = 0 & (2) \end{cases}$$

Solve (1): $\beta_0 = \bar{y} - \beta_1 \bar{x}$

plug into (2): $\sum x_i y_i - n\bar{x}(\bar{y} - \beta_1 \bar{x}) - \beta_1 \sum x_i^2 = 0$

$$\Rightarrow \beta_1 = \frac{\sum x_i y_i - n\bar{x}\bar{y}}{\sum x_i^2 - n\bar{x}^2} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2}$$

Can you prove this is true?

Recall sample variance of x is $\sum_{i=1}^{n} (x_i-\bar{x})^2/(n-1)$

sample covariance of x and y is $\sum_{i=1}^{n} (x_i-\bar{x})(y_i-\bar{y})/(n-1)$

So some people denote β_1 as $\frac{S_{xy}}{S_{xx}}$

- The LSE of β_0 and β_1 are obtained as

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

$$\hat{\beta}_1 = \frac{\sum (x_i-\bar{x})(y_i-\bar{y})}{\sum (x_i-\bar{x})^2} \quad \text{or} \quad \frac{S_{xy}}{S_{xx}}$$

Quick check for concavity:

$$\begin{cases} \dfrac{\partial^2 S}{\partial \beta_0^2} = 2n \\ \dfrac{\partial^2 S}{\partial \beta_1^2} = 2\sum x_i^2 \\ \dfrac{\partial^2 S}{\partial \beta_0 \partial \beta_1} = 2\sum x_i \end{cases} \Rightarrow \underbrace{H(\beta_0, \beta_1)}_{\text{Hessian Matrix}} = \begin{pmatrix} 2n & 2\sum x_i \\ 2\sum x_i & 2\sum x_i^2 \end{pmatrix}$$

$$\begin{aligned} \det(H) &= 4n\sum x_i^2 - 4(\sum x_i)^2 \\ &= 4n\sum x_i^2 - 4(n\bar{x})^2 = 4n\sum x_i^2 - 4n^2\bar{x}^2 \\ &= 4n(\sum x_i^2 - n\bar{x}) = 4n(\sum x_i^2 - 2n\bar{x} + n\bar{x}) \\ &= 4n(\sum x_i - 2\bar{x}\cdot\sum x_i + n\bar{x}) = 4n\sum(x_i-\bar{x})^2 > 0 \end{aligned}$$

Positive definite $\longrightarrow$ we have minimum.

IV. Mean of Variance of β_0 and β_1

How well the estimate performed can be evaluated by the mean and variance of the estimates:

- we want the estimate to be unbiased: $E(\hat{\beta}) = \beta$
- we want the estimate to be consistent: small variance.

For LSE of β_0 and β_1, we have the following theory to sum up the above:

Gauss-Markov thm:

The least-square estimates $\hat{\beta}_0$ and $\hat{\beta}_1$ are unbiased,

i.e. $E(\hat{\beta}_0) = \beta_0$, $E(\hat{\beta}_1) = \beta_1$.

Moreover, $\hat{\beta}_0$ and $\hat{\beta}_1$ are $\boxed{\text{BLUE}}$

BLUE: Best Linear Unbiased Estimator.

- Best: smallest variance among all unbiased estimators of β_0 and β_1
- Linear: the estimator is a linear function of the data.

By proving the Gauss-Markov Thm, we can get the following:

1. $E(\hat{\beta}_0) = \beta_0$

2. $E(\hat{\beta}_1) = \beta_1$

3. $Var(\hat{\beta}_0) = \sigma^2 \left(\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}} \right)$

4. $Var(\hat{\beta}_1) = \frac{\sigma^2}{S_{xx}}$

Proof. : we will start with $\hat{\beta}_1$

recall $\hat{\beta}_1 = \frac{\Sigma (x_i - \bar{x})(y_i - \bar{y})}{\Sigma (x_i - \bar{x})^2} = \frac{\Sigma (x_i - \bar{x}) y_i}{\Sigma (x_i - \bar{x})^2}$

Linear !!! $= \Sigma k_i y_i$

where $k_i = \frac{x_i - \bar{x}}{\Sigma (x_i - \bar{x})^2}$

Hint: $\Sigma (x_i - \bar{x})(y_i - \bar{y})$
$= \Sigma x_i y_i - n \bar{x} \bar{y}$
$= \Sigma (x_i - \bar{x}) y_i$

PAUSE

THREE FACTS ABOUT THE k_i's

1) $\Sigma k_i = 0$

$\Sigma k_i = \Sigma (x_i - \bar{x}) / \Sigma (x_i - \bar{x})^2 = \frac{\Sigma x_i - n\bar{x}}{\Sigma (x_i - \bar{x})^2} = \frac{n\bar{x} - n\bar{x}}{\Sigma (x_i - \bar{x})^2} = 0$

2) $\Sigma k_i^2 = \frac{1}{\Sigma (x_i - \bar{x})^2}$

$\Sigma k_i^2 = \Sigma \left(\frac{x_i - \bar{x}}{\Sigma (x_i - \bar{x})^2} \right)^2 = \frac{\Sigma (x_i - \bar{x})^2}{\left(\Sigma (x_i - \bar{x})^2 \right)^2} = \frac{1}{\Sigma (x_i - \bar{x})^2}$

37 $\Sigma k_i X_i = 1$

$$\Sigma k_i X_i = \Sigma (X_i - \bar{X}) X_i / \Sigma (X_i - \bar{X})^2 = \frac{\Sigma (X_i)^2 - n\bar{X}^2}{\Sigma (X_i - \bar{X})^2} \text{ same } = 1$$

pause ends

Now: $\hat{\beta}_1 = \Sigma k_i y_i$

$$E(\hat{\beta}_1) = E(\Sigma k_i y_i) = \Sigma k_i E(y_i) = \Sigma k_i (\beta_0 + \beta_1 X_i)$$

$$= \beta_0 \underbrace{\Sigma k_i}_{=0} + \beta_1 \underbrace{\Sigma k_i X_i}_{=1}$$

$\boxed{\text{UNBIASED!!}}$ $= \beta_1.$

$$Var(\hat{\beta}_1) = Var(\Sigma k_i y_i) = \sum_{i=1}^{n} k_i^2 Var(y_i)$$

$$= \Sigma k_i^2 \sigma^2 = \frac{\sigma^2}{\Sigma (X_i - \bar{X})^2}$$

variance in y

variance in X

S_{xx}

$\boxed{\text{Why Best? Will Answer in MLR}}$

$$E(\hat{\beta}_0) = E(\bar{y} - \bar{X}\hat{\beta}_1) = E(\bar{y}) - \bar{X} E(\hat{\beta}_1)$$

$$= \frac{1}{n} \Sigma E(y_i) - \bar{X}\beta_1$$

$$= \frac{1}{n} \Sigma (\beta_0 + \beta_1 X_i) - \bar{X}\beta_1 = \beta_0$$

$\boxed{\text{UNBIASED!!}}$

Now $\hat{\beta}_0$ is also linear, A.K.A. $\hat{\beta}_0 = \sum c_i y_i$

$$\hat{\beta}_0 = \bar{y} - \bar{x}\hat{\beta}_1 = \frac{1}{n}\sum y_i - \bar{x}\sum k_i y_i = \sum\left(\frac{1}{n} - \bar{x}k_i\right)y_i$$

$$Var(\hat{\beta}_0) = \left[\sum\left(\frac{1}{n} - \bar{x}k_i\right)^2\right]\sigma^2$$

Now, $\sum\left(\frac{1}{n} - \bar{x}k_i\right)^2$

$$= \sum\left(\frac{1}{n^2} - \frac{2\bar{x}k_i}{n} + \bar{x}^2 k_i^2\right)$$

$$\cancel{= \sum\left(\frac{1}{n^2}\right.}$$

$$= \frac{1}{n} - \frac{2\bar{x}}{n}\underbrace{\sum k_i}_{=0} + \bar{x}^2\underbrace{\sum k_i^2}_{=\frac{1}{\sum(x_i-\bar{x})^2}}$$

$$= \frac{1}{n} + \frac{\bar{x}^2}{\sum(x_i-\bar{x})^2}$$

so $Var(\hat{\beta}_0) = \sigma^2\left(\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}\right)$

Summary and take-aways of this lecture

Classic Simple Linear Regression Model

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \quad i=1, \cdots, n$$

$$E(\varepsilon_i) = 0, \quad Var(\varepsilon_i) = \sigma^2; \quad cov(\varepsilon_i, \varepsilon_j) = 0$$

$$\varepsilon_i \overset{i.i.d}{\sim} N(0, \sigma^2)$$

→

- $E(y_i) = \beta_0 + \beta_1 x_i$
- $Var(y_i) = \sigma^2$
- $y_i |_{x=x_i} \sim N(\beta_0 + \beta_1 x_i, \sigma^2)$

Least-square Estimate of β_0 and β_1

minimize $S(\beta_0, \beta_1) = \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i)^2$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

$$\hat{\beta}_1 = \frac{\Sigma (x_i - \bar{x})(y_i - \bar{y})}{\Sigma (x_i - \bar{x})^2}$$

→

- The Estimate of $E(y_i)|_{x_i}$ is given by

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$$

- The point $(\bar{x}, \bar{y})$ lies on the fitted line:

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

Gauss-Markov Thm:

The least-square estimators are the $\boxed{\text{BLUE}}$ estimators of β_0 and β_1 with

$$E(\hat{\beta}_0) = \beta_0 \quad E(\hat{\beta}_1) = \beta_1$$

$$Var(\hat{\beta}_0) = \sigma^2 \left(\frac{1}{n} + \frac{\bar{x}^2}{\Sigma (x_i - \bar{x})^2} \right)$$

$$Var(\hat{\beta}_1) = \sigma^2 / \Sigma (x_i - \bar{x})^2$$

→

- How the variance of the estimator are the smallest among linear unbiased estimators — MLR
- How to get σ^2? Next lecture