L1. Simple Linear Regression

- Overview of regression models and ML models
- SLR: model set up
- SLR: Least square Estimate of β_0 and β_1
- SLR: mean and variance of $\hat{\beta}_0$ and $\hat{\beta}_1$ (BLUE)

1. Overview

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

Regression Models	ML Models
$\boxed{\text{More about inferences}}$	$\boxed{\text{More about prediction}}$
- relationship between variables	- compare multiple approaches
- significant features	- use train and test
- inference from the data	- higher prediction accuracy when data is complex

Improve performance and deal with complexity

Foundation models and Interpretation

Strategy of Regression Modeling

statement of problem

Data collection on relevant variables

EDA: initial check

Develop one or more regression models

Model validation, criticism and selection

Identify the "best" model

Model interpretation and prediction

II. SLR : Model Set-up

1. Classic Simple Linear Regression Model

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

- y_i : i^{th} observed value of response variable Y
- x_i : i^{th} observed value of predictor / independent variable X
- n : total number of observations
- β_0 : intercept parameter
- β_1 : slope parameter
- ε_i : random error

Note :

① Classic : related to model assumptions being "standard"

② Simple : only one predictor X

③ Linear : y_i is linear to $\boxed{\text{parameters}}$ β_0 and β_1

ex : $y_i = \beta_0 + \beta_1^2/\beta_0 \, x_i + \varepsilon_i$ — 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 $E(Y_i)$ given $X = x_i$
- The goal of SLR is to estimate $E(Y_i)$

Recall:

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

- Regression Function: $E(Y_i) = E(\beta_0 + \beta_1 x_i) + E(\varepsilon_i)$

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

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

Similarly:

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

$$cov(Y_i, Y_j) = cov(\beta_0 + \beta_1 x_i + \varepsilon_i, \ \beta_0 + \beta_1 x_j + \varepsilon_j)$$

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

$$= 0$$

III. SLR: Least Square Estimate (LSE or OLS) of β_0 and β_1

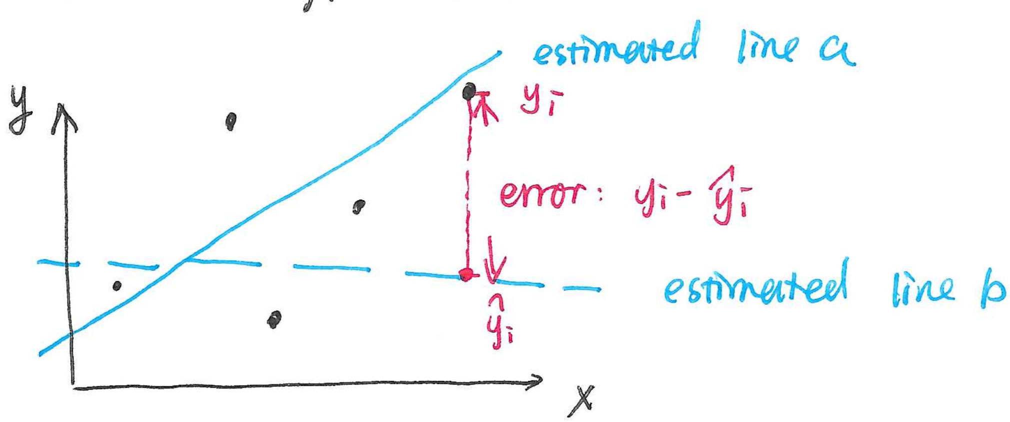
Goal: We want to estimate the linear relationship (line)

$E(Y_i) = \beta_0 + \beta_1 x_i$ between X and Y

$\Updownarrow$

Estimate β_0 and β_1

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



① Construct the deviation between the observed response y_i and the linear component:

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

② Construct a loss function to evaluate the "overall" error. The classic choice is $\boxed{\text{sum of squares}}$: $S(\beta_0, \beta_1) = \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i)^2$

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

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

Solution:

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

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

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

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

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

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

Can you prove this is true?

Recall the sample variance of X is $\frac{\sum_{i=1}^{n}(X_i-\bar{X})^2}{n-1}$

sample coviance of X and Y is $\frac{\sum_{i=1}^{n}(X_i-\bar{X})(Y_i-\bar{Y})}{n-1}$

so some peple dentote $\beta_1 = \frac{S_{xy}}{S_{xx}}$

- The OLS estimate of β_0 and β_1 are obtained as

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}$$

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(X_i-\bar{X})(Y_i-\bar{Y})}{\sum_{i=1}^{n}(X_i-\bar{X})^2} \quad \text{or} \quad \frac{S_{xy}}{S_{xx}}$$

Change of notation:

$f(x)=0 \Leftrightarrow S'(\beta_0,\beta_1)=0$

$x = x_0$ $\quad \beta_0=\hat{\beta}_0$

parameter variable $\quad$ fixed value. $\quad \beta_1=\hat{\beta}_1$

- Quick check for concavity

$$\begin{cases} \frac{\partial^2 S}{\partial \beta_0^2} = 2n \\ \frac{\partial^2 S}{\partial \beta_1^2} = 2\sum X_i^2 \\ \frac{\partial^2 S}{\partial \beta_0 \partial \beta_1} = 2\sum X_i \end{cases} \Rightarrow \underline{H(\beta_0,\beta_1)} = \begin{pmatrix} 2n & 2\sum X_i \\ 2\sum X_i & 2\sum X_i^2 \end{pmatrix}$$

Hessian Matrix

$$\det(H) = 4n\sum X_i^2 - 4(\sum X_i)^2 = 4n\sum X_i^2 - 4(n\bar{X})^2 = 4n(\sum X_i^2 - n\bar{X}^2)$$

$$= 4n(\sum X_i^2 - 2n\bar{X}^2 + n\bar{X}^2) = 4n(\sum X_i^2 - 2\bar{X}\cdot\sum X_i + n\bar{X}^2)$$

$$= 4n(X_i-\bar{X})^2 > 0$$ Positive Definite, we have mininum.

IV. Mean of Variance of $\hat{\beta}_0$ and $\hat{\beta}_1$

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 OLSE of β_0 and β_1, we have the folling 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 $\hat{\beta}_1 = \frac{\Sigma (x_i - \bar{x})(y_i - \bar{y})}{\Sigma (x_i - \bar{x})^2} = \frac{\Sigma (x_i - \bar{x}) y_i}{\Sigma (x_i - \bar{x})^2}$

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

$= \Sigma x_i y_i - n \bar{x} \bar{y}$

$= \Sigma (x_i - \bar{x}) y_i$

$\boxed{\text{Linear !!!}}$ $= \Sigma k_i y_i$

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

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

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

3) $\sum k_i x_i = 1$

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

pause ends

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

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

$$= \beta_0 \underbrace{\sum k_i}_{=0} + \beta_1 \underbrace{\sum k_i x_i}_{=1}$$

$$= \beta_1. \quad \boxed{\text{unbiased!}}$$

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

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

Variance of $\hat{\beta}_1$ is interpretated by both the variance in Y and variance in X.

- Why $Var(\hat{\beta}_1)$ is best? i.e. If B_1 is another unbiased linear estimator of β_1, then $Var(B_1) \geq Var(\hat{\beta}_1)$.

see supplemental material.

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

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

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

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

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

Summary of this lecture.

Classic Simple Linear Regression Model

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

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

$\varepsilon_i \overset{iid}{\sim} N(0, \sigma^2)$

$\longrightarrow$

$E(Y_i) = \beta_0 + \beta_1 x_i$

$Var(Y_i) = \sigma^2$

$Y_i|_{X=x_i} \sim N(\beta_0 + \beta_1 x_i, \sigma^2)$

OLSE of β_0 and β_1

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

$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}$

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

$\longrightarrow$

The estimate of $E(Y_i)$ is

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

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 OLSE are the BLUE of β_0 and β_1 with

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

$$\left.\begin{aligned} Var(\hat{\beta}_0) &= \sigma^2 \left(\frac{1}{n} + \frac{\bar{X}^2}{\Sigma (x_i - \bar{X})^2} \right) \\ Var(\hat{\beta}_1) &= \frac{\sigma^2}{\Sigma (x_i - \bar{x})^2} \end{aligned}\right\} \text{smallest among linear unbiased estimators.}$$

$\longrightarrow$

still unkown parameter: σ^2

How to estimate it?

Next lecture.