Multiple Linear Regression

outline:

- setup model & get LS est
  - ↳ matrix formulation of regression
- Gauss-Markov
- Fitted values & estimate σ^2

① Model Set Up

SLR: $Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i \qquad i = 1, \dots, n$

$$\begin{cases} Y_1 = \beta_0 + \beta_1 x_1 + \varepsilon_1 \\ Y_2 = \beta_0 + \beta_1 x_2 + \varepsilon_2 \\ \vdots \\ Y_n = \beta_0 + \beta_1 X_n + \varepsilon_n \end{cases}$$

Let's define

$$\underset{n\times 1}{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}, \quad \underset{n\times 2}{X} = \begin{bmatrix} 1 & X_1 \\ 1 & X_2 \\ \vdots & \vdots \\ 1 & X_n \end{bmatrix}, \quad \underset{2\times 1}{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}, \quad \underset{n\times 1}{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

$$Y = X\beta + \varepsilon$$

$$y = \begin{bmatrix} \beta_0 + \beta_1 x_1 \\ \beta_0 + \beta_1 x_2 \\ \vdots \\ \beta_0 + \beta_1 x_n \end{bmatrix} + \varepsilon$$

← this is equivalent to my system of n equations

Assumptions

① X is fixed

② $\varepsilon \sim N\left(0, \begin{bmatrix} \sigma^2 & 0 & 0 & \cdots & 0 \\ 0 & \sigma^2 & & & \vdots \\ \vdots & & \ddots & & \\ \vdots & & & \ddots & 0 \\ 0 & \cdots & & 0 & \sigma^2 \end{bmatrix}\right)$

$$I = \begin{bmatrix} 1 & & 0 \\ & \ddots & \\ 0 & & 1 \end{bmatrix}$$

$$= N(0, \sigma^2 I_n)$$

Aside:

if I have a random vector $Z = \begin{bmatrix} Z_1 \\ Z_2 \end{bmatrix}$

then $E(Z) = \begin{bmatrix} E(Z_1) \\ E(Z_2) \end{bmatrix} = \mu$ is the mean vector

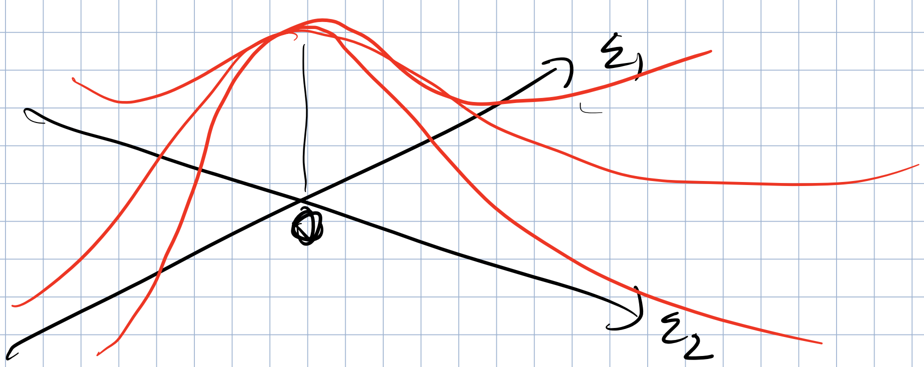
& $Var(Z) = \begin{bmatrix} Var(Z_1) & cov(Z_1, Z_2) \\ cov(Z_1, Z_2) & Var(Z_2) \end{bmatrix} = \Sigma$ is the covariance matrix.

random variable z: $\quad Var(z) = E((X-\mu)^2)$

random vector z: $\quad Var(z) = E(\underbrace{(X-\mu)}_{n\times 1}\underbrace{(X-\mu)^T}_{1\times n})$

Multivariate Normal Dist

$$\varepsilon = \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{bmatrix} \sim \underline{N(0, \sigma^2 I)}$$



$$f(\varepsilon_1, \dots, \varepsilon_n) = \prod_{i=1}^{n} f(\varepsilon_i)$$

$$= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\varepsilon_i^2/2\sigma^2}$$

$$= (2\pi\sigma^2)^{-n/2} e^{-\sum_i \varepsilon_i^2/2\sigma^2}$$

In general a random vector $Z \sim N(\mu, \Sigma)$ has the pdf:

$$f(z_1, \dots, z_n) = \frac{1}{(\sqrt{2\pi})^n |\Sigma|^{1/2}} e^{-(x-\mu)^T \Sigma^{-1} (x-\mu)/2}$$

② Parameter Estimation under Matrix Setup

In SLR the idea was to minimize

$$Q(\underline{\beta_0}, \underline{\beta_1}) = \underbrace{\sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i)^2}$$

In the matrix setup, we minimize

$$Q(\beta) = (Y - X\beta)^T (Y - X\beta)$$

$$= \| Y - X\beta \|^2$$

If I want to minimize $Q(\beta)$ w.r.t. β which is a vector, we need matrix calculus.

① $\dfrac{\partial(\beta^T A)}{\partial \beta} = A$

② $\dfrac{\partial(\beta^T A \beta)}{\partial \beta} = \underline{\underline{2A\beta}}$ if A is symmetric

③ $\dfrac{\partial(c)}{\partial \beta} = 0$

$y^T X\beta = \langle y, X\beta\rangle = \langle X\beta, y\rangle$
$= (X\beta)^T y$
$= \beta^T X^T y$

Goal is to minimize

$Q(\beta) = \|Y - X\beta\|^2 = (Y - X\beta)^T (Y - X\beta)$

$= Y^T Y - \boxed{Y^T X\beta} - (X\beta)^T Y + (X\beta)^T (X\beta)$

$= \underline{\underline{Y^T Y}} - \underset{\boxed{-\beta^T X^T Y}}{\underset{\shortparallel}{Y^T X\beta}} - \beta^T X^T Y + \underline{\beta^T X^T X\beta}$

$\hookrightarrow \dfrac{\partial Q}{\partial \beta} = 0 - X^T Y - X^T Y + 2X^T X\beta \overset{!}{=} 0$

$= -2X^T Y + 2X^T X\beta = 0$

$\Rightarrow (X^T X)\beta = X^T Y$

$$\Longrightarrow \hat{\beta} = (X^T X)^{-1} X^T Y$$ ← the Normal Eq.

$(\ldots)$ SLR

$$\begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{pmatrix} = \begin{pmatrix} \bar{Y} - \hat{\beta}_1 \bar{X} \\ \frac{SSXY}{SSX} \end{pmatrix}$$

[worked out in Cody Notes]

Why did I do this?

Now I can add more predictors to the design matrix, X:

Multiple Linear Regression Model

$$Y_i = \beta_0 + \beta_1 X_{1i} + \cdots + \beta_{p-1} X_{(p-1)i} + \varepsilon_i \qquad i = 1, \ldots, n$$

$$\Updownarrow$$

$$Y = X\beta + \varepsilon$$

where

$$\underset{n \times p}{X} = \begin{bmatrix} 1 & X_{11} & X_{21} & \cdots & X_{(p-1)1} \\ 1 & X_{12} & X_{22} & & \vdots \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & X_{1n} & X_{2n} & \cdots & X_{(p-1)n} \end{bmatrix} \quad \& \quad \beta = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_{p-1} \end{pmatrix}$$

So the LS estimate for the MLR model is just $\hat{\beta} = (X^T X)^{-1} X^T Y$ (assuming that X has full rank).

Fitted Values

$= H =$ "hat matrix"

$$\hat{Y} = X\hat{\beta} = X(X^TX)^{-1}X^TY$$

$$= HY$$