Multiple Linear Regression

Outline:

- model setup & LS estimates & their dist
  - ↳ matrix representation
- Gauss-Markov again
- fitted values & estimate σ^2

① Model Setup

SLR:

$$Y_i = \beta_0 + \beta_1 X_i + \varepsilon_i \qquad i = 1, \dots, n$$

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

In other words:

$$\left.\begin{aligned} Y_1 &= \beta_0 + \beta_1 X_1 + \varepsilon_1 \\ Y_2 &= \beta_0 + \beta_1 X_2 + \varepsilon_2 \\ &\vdots \\ Y_n &= \beta_0 + \beta_1 X_n + \varepsilon_n \end{aligned}\right\} \quad n \text{ equations}$$

Let's define $Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$

$X = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix}$, $\beta = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}$

Idea:

$$X\beta = \begin{bmatrix} \beta_0 + \beta_1 x_1 \\ \beta_0 + \beta_1 x_2 \\ \vdots \\ \beta_0 + \beta_1 x_n \end{bmatrix}$$

$\varepsilon = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}$.

Then the n equations can be succinctly written as:

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

Assumptions

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

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

Random vectors

Lets say I have a random vector $\underline{z}$

$$z = \begin{bmatrix} z_1 \\ z_2 \end{bmatrix}$$

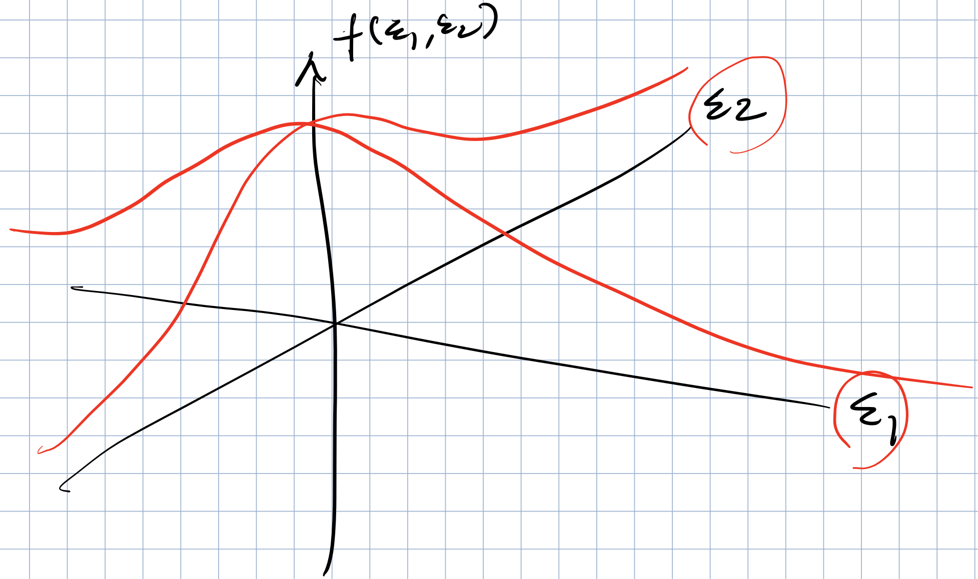
Then $\quad E(\underline{z}) = \begin{bmatrix} E(z_1) \\ E(z_2) \end{bmatrix}$ is the mean vector $= \mu$

The $\quad Var(\underline{z}) = \begin{bmatrix} Var(z_1) & Cov(z_1, z_2) \\ Cov(z_1, z_2) & Var(z_2) \end{bmatrix}$

$$= E\Big(\underset{2\times 1}{(\underline{z}-\mu)} \underset{1\times 2}{(\underline{z}-\mu)^T} \Big)$$

2×2

Multivariate Normal Dist:



$$f(\underline{\varepsilon_1}, \dots, \underline{\varepsilon_n}) = \prod_i f_i(\varepsilon_i)$$

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

$$= \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^n e^{-\sum_i \varepsilon_i^2/2\sigma^2}$$

specifically the pdf for $N(0, \sigma^2 I)$

In general for $z \sim N(\mu, \Sigma)$ the pdf is:

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

How can I estimate β?

(solving the LS problem)

Last time:

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

exact same thing

Now:

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

over all choices of β

Idea:

$$\frac{\partial Q}{\partial \beta} \overset{!}{=} 0$$

... I need matrix derivatives.

Matrix Differentiation:

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

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

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

$$\frac{\partial Q}{\partial \beta} = \frac{\partial}{\partial \beta}\left(\|Y - X\beta\|^2 \right)$$

$$= \frac{\partial}{\partial \beta}\left((Y-X\beta)^T (Y - X\beta) \right)$$

$$\beta^T X^T Y = \langle X\beta, Y\rangle = \langle Y, X\beta\rangle = Y^T X\beta$$

$$= \frac{\partial}{\partial \beta}\left(Y^T Y - Y^T X\beta - (X\beta)^T Y + (X\beta)^T (X\beta) \right)$$

$$= \frac{\partial}{\partial \beta}\left(Y^T Y - 2\,\underline{\beta^T X^T Y} + \beta^T X^T X\beta \right)$$

$$= 0 - 2(X^T Y) + 2(X^T X)\beta \overset{!}{=} 0$$

(Normal Eq)

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

↑ solve for β

$$\Rightarrow \hat{\beta} = (X^T X)^{-1} X^T Y = \underset{\beta}{\text{argmin}}\; Q(\beta)$$

If you said

$$X = \begin{bmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix} \quad \& \quad Y = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}$$

like in SLR then

$$\hat{\beta} = (X^T X)^{-1} X^T Y = \begin{bmatrix} \bar{Y} - \hat{\beta}_1 \bar{X} \\ SSXY/SSX \end{bmatrix}$$

↑ check cody Notes

Now that I have a LS solution in general, I can apply it to the case w/ multiple predictors:

Multiple Linear Regression Model

$$y_i = \underline{\beta_0} + \underline{\beta_1} X_{1i} + \underline{\beta_2} X_{2i} + \cdots + \underline{\beta_p} X_{pi} + \varepsilon_i$$

$$\Rightarrow \underbrace{X = \begin{bmatrix} 1 & X_{11} & X_{12} & \cdots & X_{1p} \\ 1 & X_{21} & X_{22} & & X_{2p} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & X_{n1} & X_{n2} & & X_{np} \end{bmatrix}}_{\text{design matrix}} \quad \& \quad \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix}.$$

From the Normal Eq:

$$\hat{\beta} = (X^T X)^{-1} X^T Y = \begin{bmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \vdots \\ \hat{\beta}_p \end{bmatrix}$$

What if I want a fitted value $\hat{Y}$?

$$\hat{Y} = X\hat{\beta} = \underbrace{X}_{n\times(p+1)}\underbrace{(X^TX)^{-1}}_{(p+1)\times(p+1)}\underbrace{X^T}_{(p+1)\times n}Y$$

$= H =$ "hat matrix"

$$= \underbrace{H}_{n\times n}\underbrace{Y}_{n\times 1}$$

Another equivalent MLR formulation:

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \cdots + \beta_{p-1} X_{(p-1)i} + \varepsilon_i$$

the col of 1s is for the intercept

$$\Rightarrow \quad X = \underbrace{\begin{bmatrix} 1 & X_{11} & X_{12} & \cdots & X_{1(p-1)} \\ 1 & X_{21} & X_{22} & & X_{2(p-1)} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & X_{n1} & X_{n2} & & X_{n(p-1)} \end{bmatrix}}_{n\times p}, \quad \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_{p-1} \end{bmatrix}$$

Still:

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

IV

$X = (\text{omnivore, vegeterian, vegan})$

$$X_1 = \mathbb{1}(\text{vegeterian}) = \begin{cases} 1 & \text{if vegeterian} \\ 0 & \text{if not} \end{cases}$$

$$X_2 = \mathbb{1}(\text{vegan})$$

EX: Write out the design matrix & the fitted vals for each cat.

$$X = \begin{bmatrix} 1 & x_{11} & x_{21} \\ 1 & x_{21} & x_{22} \\ \vdots & \vdots & \vdots \\ 1 & x_{n1} & x_{n2} \end{bmatrix} \overset{\text{ex}}{=} \begin{bmatrix} 1 & 0 & 1 \\ \vdots & 1 & 0 \\ \vdots & 0 & 0 \\ & & \vdots \\ 1 & & \end{bmatrix}$$

$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1(0) + \hat{\beta}_2(0) = \hat{\beta}_0 \leftarrow$ omnivore fitted val

$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1(1) + \hat{\beta}_2(0) = \hat{\beta}_0 + \underline{\underline{\hat{\beta}_1}} \leftarrow$ vegeterian fitted val

$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1(0) + \hat{\beta}_2(1) = \hat{\beta}_0 + \hat{\beta}_2 \leftarrow$ vegan " "

Ex:

$\hat{\beta}_0 = 170$ lbs

$\hat{\beta}_1 = -20$ lbs

$\hat{\beta}_2 = -30$ lbs