M2: Multiple linear regression [2.3, 3, B4] 27.08.2017

Notation:

independent units of observation

$i = 1, \dots, n$ ← (Y_i, x_i^T) ← p-vector

$$\begin{bmatrix} Y_1 \\ \vdots \\ Y_n \end{bmatrix} = \underset{n\times 1}{Y}$$

← random variable

vector of responses ← rent wages

fixed

$$\begin{bmatrix} x_1^T \\ x_2^T \\ \vdots \\ x_n^T \end{bmatrix} = \underset{n\times p}{X}$$

design matrix

rows = observations
cols = covariates

parameters of interest

β $(p\times 1)$ vector of regression coeff (parameter)

unknown

main focus to find β

estimate

ε: $(n\times 1)$ random vector

unknown & unobserved

Model:

$$Y = X\beta + \varepsilon \qquad \varepsilon \sim N_n(0, \sigma^2 I)$$

$$Y_i = x_i^T \beta + \varepsilon_i \qquad \varepsilon_i \sim N(0, \sigma^2)$$

$\varepsilon_i, \varepsilon_j$ independent

Other model assumptions:

Full rank: $\text{rank}(\mathbb{X}) = p$

$\underset{n\times p}{X}$ where $n >> p$

number of lin. indep. columns.

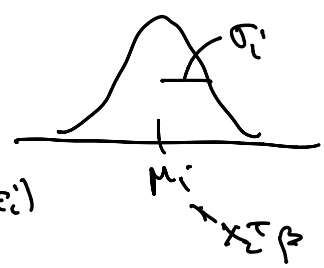
Identifiability problems when not full rank.

Distribution of Y_i

$$Y_i = x_i^T \beta + \varepsilon_i \leftarrow N(0, \sigma^2)$$

$\uparrow \beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \dots + \beta_k x_{ki}$

$Y_i \sim N(x_i^T \beta, \sigma^2)$



$E(Y_i) = E(\underline{x_i^T \beta} + \varepsilon_i) = x_i^T \beta + \underset{\overset{\shortparallel}{0}}{E(\varepsilon_i)}$

$Var(Y_i) = Var(x_i^T \beta + \varepsilon_i) = 0 + \sigma^2$

Y_i, Y_l will be independent since ε_i and ε_j is indep.

$$\underset{n\times 1}{Y} \sim N_n(\mathbb{X}\beta, \sigma^2 \mathbb{I})$$

but, we could have written

nuisance

$$Y_i \sim N(\mu_i, \sigma^2)$$ ← distribution of response

par. of interest

$$\eta_i = x_i^T \beta$$ ← linear predictor

$$\mu_i = \eta_i$$ ← connection $\underset{}{\overset{\text{mean}}{\mu_i}}$ and $\overset{\text{lin pred.}}{\eta_i}$