TMA4315 M2.w1
30.08.2018

31.08. Typo on $\cdot$ in $s(\beta)$ corrected in this colour.

Multiple linear regression ← from a GLM point of view

Notation and assumptions

$i = 1, \dots, n$ index for obs.

Response Y_i : $\underset{n\times 1}{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}$

(more on coding in IL)

Covariates (coded if categorical) and added intercept 1

$\underset{\text{obs } i}{x_i^T} = (1 \;\; x_{i1} \;\; x_{i2} \; \cdots \; x_{ik}) \Rightarrow \underset{n\times p = n\times(k+1)}{X} = \begin{bmatrix} x_1^T \\ x_2^T \\ \vdots \\ x_n^T \end{bmatrix}$

design matrix

Assume (x_i, Y_i) independent pairs.

Parameter of interest $\underset{[(k+1) = p]\times 1}{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{bmatrix}$

to be considered into a linear predictor:

$$\eta_i = x_i^T \beta = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_k \cdot x_{ik}$$

MODEL

"Traditional way":

$$\underset{n\times 1}{Y} = \underbrace{\underset{n\times p}{X} \underset{p\times 1}{\beta}}_{\begin{bmatrix} \eta_1 \\ \vdots \\ \eta_n \end{bmatrix}} + \underset{n\times 1}{\varepsilon}$$

random component called error

Assume $\underset{n}{\varepsilon} \sim N(0, \sigma^2 I)$

nuisance ϕ

identity matrix $\begin{bmatrix} 1 & 0 & \cdots & 0 \\ & 1 & \ddots & \\ 0 & & \cdots & 1 \end{bmatrix}_{n \times n}$

we will use this only for some Tzsho-like model check with residuals

↘ but more on to "GLM way"

<u>The GLM way</u> ⟵ no ε's now!

1) Random component: $Y_i \sim N(\mu_i, \sigma^2)$

not dependent on i

mean — parameter of interest

variance — nuisance

LM: exp. fam.

$$\begin{bmatrix} \theta_i = \mu_i \\ \phi = \sigma^2 \end{bmatrix}$$

2) Systematic component: $\eta_i = x_i^T \beta$

3) Link function (linking the systematic and random component)

mean of

$$\begin{tikzcd} \eta_i \arrow[r, bend left, "g"] & \mu_i \arrow[l, bend left, "h = g^{-1}"] \end{tikzcd}$$

↖ inverse function

$\eta_i = g(\mu_i)$ link function

$\mu_i = h(\eta_i)$ response (mean) function

What type of link function do we need for the MLR?

Identity link: $\eta_i = \mu_i$

Parameter estimation

Likelihood: $L(\beta) = \prod_{i=1}^{n} f_i(y_i ; \beta, \sigma^2)$

$$= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \cdot \exp\left\{-\frac{1}{2\sigma^2}(y_i - \underset{\substack{\parallel \\ \eta_i = x_i^T\beta}}{\mu_i})^2\right\}$$

$$L(\beta) = (2\pi)^{-\frac{n}{2}} (\sigma^2)^{-\frac{n}{2}} \exp\left\{-\frac{1}{2\sigma^2} \underbrace{\sum_{i=1}^{n} (y_i - x_i^T\beta)^2}_{(Y - X\beta)^T(Y - X\beta)}\right\}$$

check for yourself

Log likelihood:

$$\ell(\beta) = \ln L(\beta) = -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln(\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - x_i^T\beta)^2$$

or $\sum_{i=1}^{n}(y_i - x_i^T\beta)^2 = (Y - X\beta)^T(Y - X\beta)$

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

Score function

$$\underset{p\times 1}{s(\beta)} = \frac{\partial l}{\partial \beta} = \begin{bmatrix} \frac{\partial l}{\partial \beta_0} \\ \frac{\partial l}{\partial \beta_1} \\ \vdots \\ \frac{\partial l}{\partial \beta_k} \end{bmatrix} = \left\{ \frac{\partial l}{\partial \beta_j} \right\}$$

Alt 1:

$$\frac{\partial l(\beta)}{\partial \beta_j} = \frac{\partial}{\partial \beta_j}\left(\text{const} - \frac{1}{2\sigma^2}\sum_{i=1}^{n}\left(y_i - (\beta_0 + \beta_1 x_{i1} + \dots + \beta_j x_{ij} + \dots + \beta_k x_{ik})\right)^2\right)$$

$$= -\frac{1}{2\sigma^2}\, 2\cdot\sum_{i=1}^{n}(y_i - x_i^T\beta)(-x_{ij})$$

→ corrected after class
the derivative of
core is $(-x_{ij})$

$$\underset{p\times 1}{s(\beta)} = +\frac{1}{\sigma^2}\sum_{i=1}^{n}\underbrace{(y_i - x_i^T\beta)}_{1\times 1}\underset{p\times 1}{x_i}$$

Alt 2:

$$\frac{\partial l}{\partial \beta} = \frac{\partial}{\partial \beta}\left(\text{const} - \frac{1}{2\sigma^2}\left(Y^TY + \beta^T X^T X\beta - 2\beta^T X^T Y\right)\right]$$

$$= -\frac{1}{2\sigma^2}\left(0 + 2\cdot X^T X\beta - 2X^T Y\right)$$

$$= -\frac{1}{\sigma^2}\left(X^T X\beta - X^T Y\right) = s(\beta) \quad \left(= \frac{1}{\sigma^2}\left(X^T Y - X^T X\beta\right)\right)$$

Find Max Likelihood est (ML): $s(\hat{\beta}) = 0$

$$\mathbb{X}^T\mathbb{X}\hat{\beta} = \mathbb{X}^T Y \qquad \text{normal eq.}$$

$$\underline{\underline{\hat{\beta} = (\mathbb{X}^T\mathbb{X})^{-1}\mathbb{X}^T Y}} \qquad \leftarrow \text{closed form = analytic solution}$$

Observed Fisher information matrix : $H(\beta)$

= minus Hessian of the loglikelihood

$$H(\beta) = -\frac{\partial^2 \ell}{\partial\beta\,\partial\beta^T} = -\frac{\partial s(\beta)}{\partial \beta^T}$$

$p \times p$

$$H(\beta) = \begin{bmatrix} -\frac{\partial^2 \ell}{\partial\beta_0^2} & -\frac{\partial^2 \ell}{\partial\beta_0\partial\beta_1} & \cdots & -\frac{\partial^2 \ell}{\partial\beta_0\partial\beta_k} \\ & -\frac{\partial^2 \ell}{\partial\beta_1^2} & & \\ & & \ddots & \\ \text{symmetric} & & & -\frac{\partial^2 \ell}{\partial\beta_k^2} \end{bmatrix}$$

from above

Element: $$\frac{\partial^2 \ell}{\partial\beta_j\partial\beta_m} = \frac{\partial}{\partial\beta_m}\left((+)\frac{1}{\sigma^2}\sum_{i=1}^n (y_i - x_i^T\beta)\cdot x_{ij} \right)$$

$$= \frac{\partial}{\partial\beta_m}\left((+)\frac{1}{\sigma^2}\sum_{i=1}^n \left(y_i x_{ij} - x_i^T\beta\, x_{ij} \right)\right)$$

$\beta_0 + \beta_1 x_{i1} + \cdots + \beta_m x_{im} + \cdots$

$$= \underline{\frac{1}{\sigma^2}\sum_{i=1}^n x_{im}\cdot x_{ij}} \quad \Rightarrow \quad \underline{\underline{H(\beta) = \frac{1}{\sigma^2}\mathbb{X}^T\mathbb{X}}}$$

See below - added after class.

Expected Fisher information matrix

$$F(\beta) = E(H(\beta))$$

$$F(\beta) = E(H(\beta)) = E\left(\frac{1}{\sigma^2} X^T X\right) = \frac{1}{\sigma^2} X^T X$$

no random elements

6

Added after class

Alt:
$$H(\beta) = -\frac{\partial^2 l}{\partial\beta\partial\beta^T} = -\frac{\partial}{\partial\beta^T}\left[-\frac{1}{\sigma^2}\left(X^T X\beta - X^T Y\right)\right]$$

$$= \frac{1}{\sigma^2} X^T X$$

since

$$\frac{\partial^2 l}{\partial\beta\partial\beta^T} = \frac{\partial}{\partial\beta\partial\beta^T}\left[-\frac{1}{2\sigma^2}\left(Y^T Y - \beta^T X^T X\beta - 2\beta^T X^T Y\right)\right.$$

$$= \dots = \frac{\partial}{\partial\beta\partial\beta^T}\left[+\frac{1}{2\sigma^2}\beta^T X^T X\beta\right] = \frac{1}{\sigma^2} X^T X$$

using rule 3: $\frac{\partial^2 \beta^T D\beta}{\partial\beta\partial\beta^T} = 2D$