

9-5-2025

R-lab : Splines

lm function

$\text{lm} (\underbrace{y \sim f(x)}_{\substack{\text{formula} \\ \text{object}}}, \dots, \text{data}, \dots)$

$\frac{y \sim x}{\updownarrow} : \text{formula (stat. model)}$

$$y = b_0 + b_1 \cdot x$$

$$y \sim x_1 + x_2 \Leftrightarrow y = b_0 + b_1 x_1 + b_2 x_2$$

$$y \sim x, \quad x: \text{factor levels} \begin{cases} l_1 \\ l_2 \\ l_3 \end{cases} \Leftrightarrow y = b_0 + b_1 \cdot 1(x=l_1) + b_2 \cdot 1(x=l_2)$$

$$y \sim x_1 + x_2 + \underbrace{x_1 * x_2}_{\text{interaction}}$$

$$y \sim \exp(x) \quad \times \quad \Leftrightarrow y = b_0 + b_1 e^x$$

$$y \sim \underbrace{\text{I}(\exp(x))}_{\substack{\uparrow \\ \text{modifier}}} \rightarrow \text{apply. logarithm to } e^x$$

$$y = b_0 + b_1 x + b_2 x^2 \Leftrightarrow y \sim x + I(x^2)$$

$\uparrow$
 n

$$y \sim \text{poly}(x, 2, \text{raw}=\text{TRUE})$$

(raw=FALSE - default
 unordered orthogonal polynomials)

$$y = b_0 + b_1 g_1(x) + b_2 g_2(x)$$

g_1, g_2 not bad. 2
 orthogonal

Null model: $y = b_0$ ($\Leftrightarrow E(y) = E(\bar{y})$)

$$\Leftrightarrow y \sim 1$$

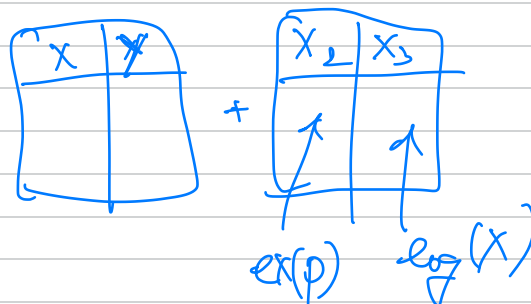
$$y \sim . \Leftrightarrow y = b_0 + b_1 x_1 + \dots + b_p x_p \quad \left(\begin{array}{l} \text{of } \\ b_0, p=2 \\ \text{on dataset} \end{array} \right)$$

$$y \sim . - X^2 \quad y = b_0 + b_1 x_1 + b_2 x_2 + \dots + b_p x_p$$

Παραδ. $y = b_0 + b_1 x + b_2 e^x + b_3 \log x$

α $y \sim x + I(\exp(x)) + I(\log(x))$

β Σ20 dataset



$$y \sim x + x_2 + x_3$$

$$\text{poly}(x, z) =$$

⑧ Basic Splines

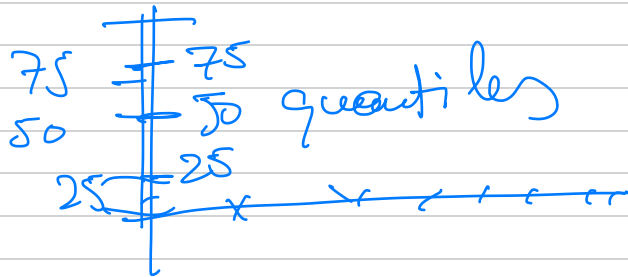
$bs(x, \text{knots} = c(10, 20, 30))$, degree df

$1 \quad x \quad x^2 \quad x^3 \quad (x-10)^3 \quad (x-20)^3 \quad (x-30)^3$

$\Downarrow$
 zu Normalform

$$bs(x, df=7)$$

$$df = 7 \Rightarrow K = 3 \quad \underline{\underline{3 \text{ Köpfe!}}}$$



$$x = (10, 10, 20, 30), \quad k=2$$



ns : natural splines

Smoothing :

smooth.spline(x, y, df)

Γοδοιναμοι βαρφοι εφειδεριας
df(λ).

smooth.spline(x, y, cv=TRUE)

↓
df* (Typei cov
df no max
to cross
validation error)