Foreløbig model

$$\underset{\nearrow}{y} = \underbrace{r(x_1, \dots, x_p)} + u \nwarrow$$

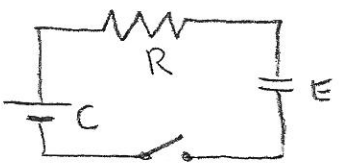
obs. værdi af stok. var.

forklarende variable (ikke stokastiske)

stokastisk led (ikke direkte observerbart)

bemærk $\underbrace{\text{sum af systematisk og stokastisk led}}_{\text{regressionsmodel}}$

eks.



↑
slutter til tiden $t = 0$

$$v + Ri = E \quad , \quad i = Cv'$$

$$v' + \frac{1}{RC} v = \frac{E}{RC} \quad , \quad RC = T$$

hom. løsn.: $\frac{v'}{v} = -\frac{1}{T}$

$$\Rightarrow \ln v = -\frac{t}{T} + k_1$$

$$\Rightarrow v = k e^{-\frac{t}{T}}$$

part. løsn.: $v_p = E$ (ses ved indsættelse)

beg. bet. $v(0) = 0 \Rightarrow k = -v_p(0) = -E$

altså $v = E(1 - e^{-\frac{t}{T}})$, $T = RC$, $t \geq 0$

mere realistisk: $v = E(1 - e^{-\frac{t}{T}}) + u$

E og T forkl. var.

bemærk ikke-lineært udtryk

ved Taylorudvikling om $t = t_0$ kan fås

$$v \approx \underbrace{E(1 - e^{-\frac{t_0}{T}})}_{\beta_0} + \underbrace{\frac{E}{T} e^{-\frac{t_0}{T}}}_{\beta_1} \underbrace{(t - t_0)}_{x} - \underbrace{\frac{E}{2T^2} e^{-\frac{t_0}{T}}}_{\beta_2} \underbrace{(t - t_0)^2}_{x^2}$$

lineært i β_0, β_1 og β_2

Når $r(\cdot)$ er lineær i $\beta_1, \dots, \beta_p$ bliver y udtrykt

$$y = \beta_1 x_1 + \dots + \beta_p x_p + u,$$

hvor $\beta_1, \dots, \beta_p$ er ukendte parametre

Egenskaber ved modellen

- simpel
- kan ofte argumenteres ud fra erfaring
- modellen kan i visse tilfælde frem-komme efter transformation, fx

$$y = ab^x \Rightarrow \ln y = \ln a + x \ln b \Leftrightarrow z = \alpha + \beta x$$

- kun linearitet i parametre nødvendig
- evt. 'linearisering' lokalt ved Taylor-udvikling
- simpel at behandle matematisk
- både kvantitative og kvalitative størrelser kan optræde som for-klarende variable

Den lineære model

$$Y_i = \beta_1 x_{i1} + \dots + \beta_p x_{ip} + U_i, \quad i = 1, \dots, n$$

uafh.

x_{ik} er i'te værdi af den k'te forkl. var.

kort notation: $Y = X\beta + U$

kaldes designmatrix

format $n \times p$, $n \geq p$

Krav: $E[U] = 0$
$\mathrm{Var}[U] = \sigma^2 I$ $\Big\}$ anden ordens hypotesen

Heraf $E[Y] = X\beta$
$\mathrm{Var}[Y] = \sigma^2 I$

Ønske: At bestemme β, så $\|y-\mu\|$ minimeres ($\mu = X\beta$)

Ækvivalent: At bestemme β, så $\|y-\mu\|^2$ minimeres

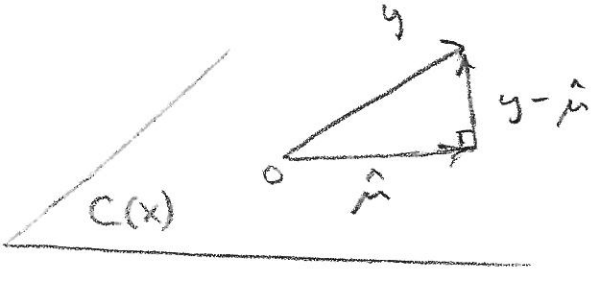
Bemærk: $\|y-\mu\|^2 = (y-\mu)^T(y-\mu)$
$= (y-X\beta)^T(y-X\beta)$

Metoden kaldes mindste kvadraters metode

Geometrisk bestemmelse

Bemærk $X\beta = \beta_1 x_1 + \dots + \beta_p x_p$

dvs. $X\beta \in \mathrm{sp}\{x_1, \dots, x_p\} = C(X)$
$\uparrow$
p-dim. underrum af R^n, når rang $X = p$



$y \in R^n$
$\hat{\mu} \in C(X)$
$y - \hat{\mu} \in (C(X))^{\perp}$

$$(y-\hat{\mu})^T X = 0 \Rightarrow (y - X\hat{\beta})^T X = 0$$

$$\Leftrightarrow X^T(y - X\hat{\beta}) = 0$$

$$\Leftrightarrow \underbrace{X^T X \hat{\beta} = X^T y}$$

kaldes normalligningerne

med rang $X = p$ får vi entydig løsning:

$$\hat{\beta} = (X^T X)^{-1} X^T y$$

heraf $\hat{\mu} = X\hat{\beta} = X(X^T X)^{-1} X^T y = Py$

$P = X(X^T X)^{-1} X^T$ er projektionsmatrix for ortogonalprojektion på $C(X)$.

Bemærk, at

- $P^T = P$ (P sym.)
- $P^2 = P$ (P idempotent)
- $\operatorname{tr} P = \operatorname{tr}(X(X^T X)^{-1} X^T)$

$$= \operatorname{tr}(X^T X (X^T X)^{-1})$$

$$= \operatorname{tr} I_p$$

$$= p$$

dvs. rang $P = \operatorname{tr} P$

Residualvektoren

$$y - \hat{\mu} = y - Py = (I_n - P)\,y$$

$I_n - P$ er projektionsmatrix for projektion på $(C(X))^{\perp}$

$y = \hat{\mu} + (y - \hat{\mu})$ (entydig dekomposition)

$$\Rightarrow \|y\|^2 = \|\hat{\mu}\|^2 + \|y - \hat{\mu}\|^2$$

(jf. Pythagoras)

Analytisk bestemmelse

$$\frac{d}{d\beta}(y - X\beta)^T(y - X\beta)$$

$$= \frac{d}{d\beta}\left(y^T y - 2y^T X\beta + \beta^T X^T X\beta\right)$$

$$= -2X^T y + 2X^T X\beta = 2(X^T X\beta - X^T y)$$

$= 0$ for $X^T X\beta = X^T y$

dvs. normalligningerne som før

$$\frac{d^2}{d\beta\, d\beta^T}(y - X\beta)^T(y - X\beta) = 2X^T X > 0$$

(pos. def.)

dvs. minimum

(bemærk, at $\frac{d}{dx}Ax = A^T$ og $\frac{d}{dx}x^T Bx = 2Bx$)

Mindste kvadraters estimator

Vi har $\hat{\beta} = (X^T X)^{-1} X^T Y$

$$\Rightarrow E[\hat{\beta}] = (X^T X)^{-1} X^T E[Y]$$
$$= (X^T X)^{-1} X^T X \beta$$
$$= \beta,$$

dvs. $\hat{\beta}$ er et centralt estimat for β.

$$\mathrm{Var}[\hat{\beta}] = (X^T X)^{-1} X^T \mathrm{Var}[Y] X (X^T X)^{-1}$$
$$= (X^T X)^{-1} X^T \sigma^2 I X (X^T X)^{-1}$$
$$= \sigma^2 (X^T X)^{-1}$$

Heraf $E[\hat{\mu}] = X E[\hat{\beta}] = X\beta = \mu$

$$\mathrm{Var}[\hat{\mu}] = X \mathrm{Var}\hat{\beta} X^T$$
$$= \sigma^2 X (X^T X)^{-1} X^T$$
$$= \sigma^2 P$$

Krav til konsistens af $\hat{\beta}$:

$((X^T X)^{-1})_{ii} \to 0$ for $n \to \infty$

bemærk $X^T X = \sum_{i=1}^{n} \tilde{x}_i \tilde{x}_i^T$ *

dvs. $(X^T X)^{-1} \to 0$ for $n \to \infty$

medmindre $\tilde{x}_i \to 0$ 'hurtigt'

(ikke-præcist argument)

* $\tilde{x}_i$ består af komponenterne i i'te rækkevektor i X

Estimat for σ^2

$$Var[U_i] = \sigma^2 \quad \Rightarrow \quad E[U_i^2] = \sigma^2$$

derfor vælges

$$\hat{\sigma}^2 = \frac{1}{n} \| y - \hat{\mu} \|^2$$ ($y - \hat{\mu}$ er residual-vektoren)

$$\begin{aligned} n\hat{\sigma}^2 &= (y-\hat{\mu})^T (y-\hat{\mu}) \\ &= ((I-P)y)^T (I-P)y \\ &= y^T (I-P) y \\ &= u^T (I-P) u \qquad * \\ &= y^T (y - X\hat{\beta}) \end{aligned}$$

$*$ idet $(I-P)\mu = \mu - P\mu = \mu - \mu = 0$

$$\begin{aligned} E[n\hat{\sigma}^2] &= E[U^T (I-P) U] \\ &= 0^T (I-P) 0 + tr((I-P) \, Var\, U) \\ &= 0 + tr(\sigma^2 (I-P)) \\ &= (n-p)\sigma^2 \end{aligned}$$

dvs. $s^2 = \frac{n}{n-p} \hat{\sigma}^2$ er et centralt estimat for σ^2

eks. $\quad y = 1_n \beta + u \quad$ (n uafh. obs.)

$$(1_n^T 1_n)\beta = 1_n^T y \quad \Leftrightarrow \quad n\beta = \sum_i y_i$$

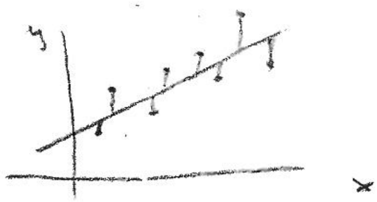
$$\Rightarrow \hat{\beta} = \frac{n\bar{y}}{n} = \bar{y} \quad \wedge \quad Var\, \hat{\beta} = \sigma^2 (1_n^T 1_n)^{-1}$$

$$= \frac{\sigma^2}{n}$$

eks. $y = \beta_0 1_n + \beta_1 x + u$ (n obs.)

(simpel lineær regresjon)

$$X = \begin{bmatrix} 1 & x_1 \\ \vdots & \vdots \\ 1 & x_n \end{bmatrix} \quad n \times 2$$



$$X^T X = \begin{bmatrix} n & \Sigma x_i \\ \Sigma x_i & \Sigma x_i^2 \end{bmatrix} = \begin{bmatrix} n & n\bar{x} \\ n\bar{x} & \Sigma x_i^2 \end{bmatrix}$$

$$(X^T X)^{-1} = \frac{1}{n \Sigma x_i^2 - n^2 \bar{x}^2} \begin{bmatrix} \Sigma x_i^2 & -n\bar{x} \\ -n\bar{x} & n \end{bmatrix}$$

$$X^T y = \begin{bmatrix} \Sigma y_i \\ \Sigma x_i y_i \end{bmatrix} = \begin{bmatrix} n\bar{y} \\ \Sigma x_i y_i \end{bmatrix}$$

$$\hat{\beta} = \frac{1}{n S_{xx}} \begin{bmatrix} \Sigma x_i^2 & -n\bar{x} \\ -n\bar{x} & n \end{bmatrix} \begin{bmatrix} n\bar{y} \\ \Sigma x_i y_i \end{bmatrix}$$

$$= \frac{1}{n S_{xx}} \begin{bmatrix} n\bar{y}\Sigma x_i^2 - n\bar{x}\Sigma x_i y_i \\ -n^2\bar{x}\bar{y} + n\Sigma x_i y_i \end{bmatrix}$$

$$= \frac{1}{S_{xx}} \begin{bmatrix} \bar{y}(\Sigma x_i^2 - n\bar{x}^2) - \bar{x}\Sigma x_i y_i + n\bar{x}^2\bar{y} \\ \Sigma x_i y_i - n\bar{x}\bar{y} \end{bmatrix}$$

$$= \frac{1}{S_{xx}} \begin{bmatrix} \bar{y} S_{xx} - \bar{x} S_{xy} \\ S_{xy} \end{bmatrix} = \begin{bmatrix} \bar{y} - \frac{S_{xy}}{S_{xx}}\bar{x} \\ \frac{S_{xy}}{S_{xx}} \end{bmatrix}$$

dvs. $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$ og $\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}}$

evt. forenkling:

$$z = x - \bar{x} 1_n \Rightarrow \bar{z} = \bar{x} - \bar{x} = 0$$

$$y = \alpha 1_n + \beta_2 z + u, \quad \alpha = \beta_0 + \beta_1 \bar{x}$$

$\hat{\alpha} = \bar{y}$, $\hat{\beta}_1 = \frac{S_{xy}}{S_{xx}}$, $\hat{\alpha}$ og $\hat{\beta}_1$ ukorrelerede

Stikprøvekorrelation

$$r = \frac{S_{xy}}{\sqrt{S_{xx} S_{yy}}} = \frac{\frac{1}{n-1} S_{xy}}{\sqrt{\frac{1}{n-1} S_{xx} \frac{1}{n-1} S_{yy}}} = \frac{s_{xy}}{s_x s_y}$$

Determinationskoefficienten

$$r^2 = \frac{\| \hat{y} - \bar{y} 1_n \|^2}{\| y - \bar{y} 1_n \|^2} = \frac{\| (\bar{y} - \frac{S_{xy}}{S_{xx}} \bar{x}) 1_n + \frac{S_{xy}}{S_{xx}} x - \bar{y} 1_n \|}{n S_{yy}}$$

$$= \frac{\| \frac{S_{xy}}{S_{xx}} (x - \bar{x} 1_n) \|^2}{n S_{yy}} = \frac{\left(\frac{S_{xy}}{S_{xx}}\right)^2 n S_{xx}}{n S_{yy}}$$

$$= \frac{S_{xy}^2}{S_{xx} S_{yy}}$$

(r^2 er den brøkdel af variationen, der kan forklares af modellen)

Gauss - Markov

Vi har $\hat{\beta} = (X^T X)^{-1} X^T Y$ med $E[\hat{\beta}] = \beta$

Antag $T = C^T Y$, hvor $E[T] = \beta$ for alle β

(C er $n \times p$)

Gauss-Markovs sætn.: $Var[\hat{\beta}] \leq Var[T]$

Bevis: $Var[\hat{\beta}] = \sigma^2 (X^T X)^{-1}$

$Var[T] = C^T \sigma^2 I \, C = \sigma^2 C^T C$

bemærk $E[C^T Y] = \beta \Rightarrow C^T E[Y] = \beta$

$\Rightarrow C^T X \beta = \beta \Rightarrow C^T X = I_p$,

dvs. $Var[\hat{\beta}]$ kan skrives

$$Var[\hat{\beta}] = \sigma^2 C^T X (X^T X)^{-1} X^T C$$
$$= \sigma^2 C^T P C$$

$$Var[T] - Var[\hat{\beta}] = \sigma^2 C^T C - \sigma^2 C^T P C$$
$$= \sigma^2 C^T (I - P) C \geq 0$$

(pos. semidefinit)

Vælges $\psi = a^T \beta$, er $a^T \hat{\beta}$ et centralt estimat for ψ, ifølge Gauss-Markow det centrale estimat med mindst varians.

Specielt for $a = e_j$ får vi $e_j^T \beta = \beta_j$. $\hat{\beta}_j$ er derfor bedste lineære estimat for β_j.

Middelkvadrater

Tilbage til $\|y\|^2 = \|\hat{\mu}\|^2 + \|y - \hat{\mu}\|^2$, som også kan skrives $y^T y = y^T P y + y^T (I - P) y$, hvor alle tre led er kvadratiske former (idet $y^T y = y^T I y$).

Bemark rang $I = n$, rang $P = p$, rang $(I-P) = n-p$

De såkaldte middelkvadrater defineres som

$$\frac{1}{n} y^T y \,, \quad \frac{1}{p} y^T P y \,, \quad \frac{1}{n-p} y^T (I-P) y$$

De tilsvarende stokastiske udtryk er

$$\frac{1}{n} Y^T Y \,, \quad \frac{1}{p} Y^T P Y \,, \quad \frac{1}{n-p} Y^T (I-P) Y$$

Middelværdierne af disse er

$$E\left[\frac{1}{n} Y^T Y\right] = \frac{1}{n}\left(\sigma^2 \operatorname{tr} I + \mu^T \mu\right) = \sigma^2 + \frac{1}{n} \mu^T \mu$$

$$E\left[\frac{1}{p} Y^T P Y\right] = \frac{1}{p}\left(\sigma^2 \operatorname{tr} P + \mu^T P \mu\right)$$

$$= \sigma^2 + \frac{1}{p} \mu^T P \mu$$

$$E\left[\frac{1}{n-p} Y^T (I-P) Y\right] = \frac{1}{n-p} \sigma^2 \operatorname{tr}(I-p) = \sigma^2$$