

TEKNINEN YHTEENVETO 2

Malliavusteinen estimointi

Regressioestimointi

Oletetaan lineaarinen regressiomalli (superpopulaatiomalli)

$$y_k = \alpha + \beta z_k + \varepsilon_k, \quad V(y_k) = \sigma^2 \text{ (varianssi)}$$

Äärellisen perusjoukon vastineet parametreille α ja β ovat A ja B

A ja B estimoidaan painotetulla PNS-menetelmällä (WLS)

SRS-tilanne:

Kulmakertoimen B estimaattori $\hat{b} = \hat{s}_{yz} / \hat{s}_z^2$

Vakioparametrin A estimaattori $\hat{a} = \bar{y} - b\bar{z}$

Kokonaismäärän T regressioestimaattori:

$$\hat{t}_{reg} = N(\bar{y} + \hat{b}(\bar{Z} - \bar{z})) = \hat{t} + \hat{b}(T_z - \hat{t}_z), \text{ missä}$$

$\hat{t} = N\bar{y}$ on tulosmuuttujan y kokonaismäärän T SRS-estimaattori,

$\hat{t}_z = N\bar{z}$ on apumuuttujan z kokonaismäärän T_z SRS-estimaattori

$$\bar{Z} = T_z / N$$

Tarvittava lisätieto: Apumuuttujan kokonaismäärä T_z

Asetelmavarianssi (likimääräinen)

$$V_{SRS}(\hat{t}_{reg}) \cong N^2 \left(1 - \frac{n}{N}\right) \left(\frac{1}{n}\right) S_y^2 (1 - \rho_{yz}^2),$$

missä $\rho_{yz} = S_{yz} / S_y S_z$ on muuttujien y ja z perusjoukon korrelaatio

Varianssiestimaattori (yksi vaihtoehto)

$$\hat{v}_{SRS}(\hat{t}_{reg}) = N^2 \left(1 - \frac{n}{N}\right) \left(\frac{1}{n}\right) \hat{s}_y^2 (1 - \hat{\rho}_{yz}^2)$$

Monimuuttujainen regressiomalli

$$y_k = \beta_0 + \beta_1 z_{1k} + \beta_2 z_{2k} + \dots + \beta_p z_{pk} + \varepsilon_k = \mathbf{z}'_k \boldsymbol{\beta} + \varepsilon_k$$

missä

$$\mathbf{z}_k = (1, z_{1k}, \dots, z_{pk})'$$

$$\boldsymbol{\beta} = (\beta_0, \beta_1, \dots, \beta_p)'$$

Estimaattorit ja varianssiestimaattorit

Perusmuoto

$$\hat{t}_{reg} = \hat{t}_{HT} + \hat{b}_1 (T_{z_1} - \hat{t}_{z_1}) + \hat{b}_2 (T_{z_2} - \hat{t}_{z_2}) + \dots + \hat{b}_p (T_{z_p} - \hat{t}_{z_p}) \quad (1)$$

PROC SURVEYREG-muoto

$$\hat{t}_{reg} = \hat{b}_0 N + \hat{b}_1 T_{z_1} + \hat{b}_2 T_{z_2} + \dots + \hat{b}_p T_{z_p} \quad (2)$$

GREG-muoto

$$\hat{t}_{reg} = \sum_{k=1}^N \hat{y}_k + \sum_{k=1}^n w_k (y_k - \hat{y}_k) \quad (3)$$

missä $\hat{y}_k = \mathbf{z}'_k \hat{\mathbf{b}}$

Kalibrointiestimaattori

$$\hat{t}_{reg} = \sum_{k=1}^n w_k^* y_k \quad (4)$$

missä $w_k^* = g_k w_k$ ja g_k on g-paino alkiolle k

Varianssiestimaattorit

$$\hat{v}(\hat{t}_{reg}) = N^2(1 - n/N)(1/n)\hat{s}_{\hat{e}}^2 \quad (5)$$

missä $\hat{s}_{\hat{e}}^2 = \sum_{k=1}^n (\hat{e}_k - \bar{\hat{e}})^2 / (n-1)$

missä

$$\hat{e}_k = y_k - \hat{y}_k$$

$$\bar{\hat{e}} = \sum_{k=1}^n \hat{e}_k / n$$

$$\hat{y}_k = \mathbf{z}'_k \hat{\mathbf{b}}$$

$$\hat{v}(\hat{t}_{reg}) = N^2(1 - n/N)(1/n)(n-1)/(n-p)\hat{s}_{\hat{e}^*}^2 \quad (6)$$

missä

$$\hat{s}_{\hat{e}^*}^2 = \sum_{k=1}^n (\hat{e}_k^* - \bar{\hat{e}}^*)^2 / (n-1)$$

$$\hat{e}_k^* = g_k \hat{e}_k \text{ (g-painotetut jäännökset) ja } \bar{\hat{e}}^* = \sum_{k=1}^n \hat{e}_k^* / n$$

$$\hat{v}(\hat{t}_{reg}) = N^2(1 - n/N)(1/n)\hat{s}_y^2(1 - \hat{R}^2) \quad (7)$$

missä

\hat{R}^2 on yhteiskorrelaatiokertoimen neliö

$$\hat{s}_y^2 = \sum_{k=1}^n (y_k - \bar{y})^2 / (n-1)$$

Kalibrointimenetelmä (Calibration)

g-painot

Tulosmuuttuja y

Kalibrointiestimaattori

$$\hat{t}_{reg} = \sum_{k=1}^n w_k^* y_k,$$

missä $w_k^* = g_k w_k$, g_k on g-paino alkiolle k ja $w_k = 1/\pi_k$

jolle pätee $\hat{t}_{reg} = \sum_{k=1}^n w_k^* z_k = \sum_{k=1}^N Z_k = T_z$ (kalibrointiominaisuus)

Suhde-estimointi (Ratio estimation)

Jatkuva apumuuttuja z

$$g_k = \frac{T_z}{\hat{t}_z}$$

Regressioestimointi (Regression estimation)

Jatkuva apumuuttuja z

$$g_k = \frac{N}{\hat{N}} \left(1 + \frac{\bar{Z} - \bar{z}}{(n-1)\hat{s}_z} (z_k - \bar{z}) \right)$$

n

Jälkiositus (Poststratification)

Diskreetti apumuuttuja (luokat $g = 1, \dots, G$)

$$g_{gk} = \frac{N_g}{\hat{N}_g}, \quad g = 1, \dots, G$$

missä $\hat{N}_g = \sum_{k=1}^{n_g} w_{gk}$ on jälkiositteen g estimoitu koko