

Udeladelse af én forklarende variabel

Kommentar til side 55 og 96

I det følgende studeres effekten af, at en forklarende variabel udelades fra en regressionsmodel med p forklarende variable. For nemheds skyld antages, at den p 'te, altså den sidste, udelades. De observerede værdier af denne variabel er udtrykt ved $n \times 1$ søjlen $\mathbf{X}_p$ mens de resterende søjler, dvs. de første $p - 1$ søjler i designmatricen $\mathbf{X}$, betegnes $\mathbf{X}_{(p)}$. Da gælder

$$\mathbf{X}^T \mathbf{X} = \begin{pmatrix} \mathbf{X}_{(p)}^T \mathbf{X}_{(p)} & \mathbf{X}_{(p)}^T \mathbf{X}_p \\ \mathbf{X}_p^T \mathbf{X}_{(p)} & \mathbf{X}_p^T \mathbf{X}_p \end{pmatrix}.$$

Med notationen fra lærebogen side 304 øverst, hvor den inverse af en blokmatrix er angivet, gælder

$$\mathbf{A} = \mathbf{X}_{(p)}^T \mathbf{X}_{(p)}$$

$$\mathbf{B} = \mathbf{X}_{(p)}^T \mathbf{X}_p$$

$$\mathbf{B}^T = \mathbf{X}_p^T \mathbf{X}_{(p)}$$

$$\mathbf{D} = \mathbf{X}_p^T \mathbf{X}_p$$

og

$$\begin{aligned} \mathbf{E} &= \mathbf{X}_p^T \mathbf{X}_p - \mathbf{X}_p^T \mathbf{X}_{(p)} (\mathbf{X}_{(p)}^T \mathbf{X}_{(p)})^{-1} \mathbf{X}_{(p)}^T \mathbf{X}_p \\ &= \mathbf{X}_p^T (\mathbf{I} - \mathbf{H}_{(p)}) \mathbf{X}_p, \end{aligned}$$

hvor

$$\mathbf{H}_{(p)} = \mathbf{X}_{(p)} (\mathbf{X}_{(p)}^T \mathbf{X}_{(p)})^{-1} \mathbf{X}_{(p)}^T$$

er hatmatricen i regressionsmodellen, hvori kun de $p - 1$ første forklarende variable anvendes. Desuden er

$$\mathbf{F} = (\mathbf{X}_{(p)}^T \mathbf{X}_{(p)})^{-1} \mathbf{X}_{(p)}^T \mathbf{X}_p.$$

Estimatet β_p for koefficienten til den p 'te forklarende variabel er det p 'te element i vektoren $\boldsymbol{\beta} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}$. Det findes ved hjælp af $(\mathbf{X}^T \mathbf{X})^{-1}$ og $\mathbf{X}^T$ opdelt i blokmatricer ved

$$\begin{aligned} \beta_p &= -\mathbf{E}^{-1} \mathbf{F}^T \mathbf{X}_{(p)}^T \mathbf{y} + \mathbf{E}^{-1} \mathbf{X}_p^T \mathbf{y} = \mathbf{E}^{-1} [-\mathbf{X}_p^T \mathbf{X}_{(p)} (\mathbf{X}_{(p)}^T \mathbf{X}_{(p)})^{-1} \mathbf{X}_{(p)}^T + \mathbf{X}_p^T] \mathbf{y} \\ &= \mathbf{E}^{-1} \mathbf{X}_p^T (\mathbf{I} - \mathbf{H}_{(p)}) \mathbf{y}. \end{aligned}$$

Faktoren $(\mathbf{I} - \mathbf{H}_{(p)}) \mathbf{y}$ i denne formel er residualerne i regressionsmodellen, hvori kun de $p - 1$ første forklarende variable anvendes. Disse residualer betegnes i lærebogen side 55 og 96

$$\hat{\mathbf{u}}_p = (\mathbf{I} - \mathbf{H}_{(p)}) \mathbf{y}.$$

Tilsvarende betegnes side 55 og 96

$$\hat{\mathbf{v}}_p = (\mathbf{I} - \mathbf{H}_{(p)}) \mathbf{X}_p$$

residualerne i en regressionsmodel med de første $p - 1$ forklarende variable, og hvori den p 'te forklarende variabel anvendes som responsvariabel. Da matricen $\mathbf{E}$ kan skrives som

$$\mathbf{E} = \mathbf{X}_p^T (\mathbf{I} - \mathbf{H}_{(p)})^T (\mathbf{I} - \mathbf{H}_{(p)}) \mathbf{X}_p = \hat{\mathbf{v}}_p^T \hat{\mathbf{v}}_p,$$

fordi $\mathbf{H}_{(p)}$ er idempotent og symmetrisk, ses at

$$\beta_p = (\hat{\mathbf{v}}_p^T \hat{\mathbf{v}}_p)^{-1} \hat{\mathbf{v}}_p^T \hat{\mathbf{u}}_p,$$

altså at β er MK-estimatoren i regressionsmodellen uden intercept med responsvariabel $\hat{\mathbf{u}}_p$ og $\hat{\mathbf{v}}_p$ som den eneste forklarende variabel. Dette resultat anvendes side 55 og det danner baggrund for de tilføjede variabel-diagrammer, som indføres side 96.