

Inferens i en enkel lineær regresjonsmodell

TMA4240/TMA4245 Statistikk

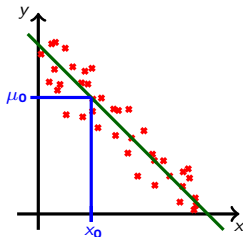
Håkon Tjelmeland

Institutt for matematiske fag

Norges teknisk-naturvitenskapelige universitet

Pivotal og testobservator

- ★ Husk: For å utlede et konfidensintervall for θ trenger vi en pivotal: $Z = h(Y_1, Y_2, \dots, Y_n, \theta)$
 - inneholder ingen ukjente parametre unntatt θ
 - har en kjent fordeling (som ikke avhenger av θ)
- ★ Husk: I en hypotesetest trenger vi en testobservator: T
 - skal være en observator: kan ikke inneholde ukjente parametre
 - må ha en kjent fordeling når H_0 er sann
- ★ Fokus i denne videoen: I en enkel lineær regresjonsmodell: $Y_i = \alpha + \beta x_i + \varepsilon_i$
 - hvilken pivotal/testobservator skal vi bruke hvis vi er interessert i α (når σ^2 ukjent)?
 - hvilken pivotal/testobservator skal vi bruke hvis vi er interessert i β (når σ^2 ukjent)?
 - hvilken pivotal for å utlede konfidensintervall for $\mu_0 = E[Y|x_0] = \alpha + \beta x_0$ (når σ^2 ukjent)?



- ★ Starter med å minne om noen egenskaper til $\hat{\alpha}$, $\hat{\beta}$ og S^2 vi allerede har funnet

Inferens om konstantleddet α

- ★ Har tidligere funnet at

$$\hat{\alpha} \sim N\left(\alpha, \sigma^2 \cdot \frac{\sum_{i=1}^n x_i^2}{n \sum_{i=1}^n (x_i - \bar{x})^2}\right), \quad \frac{(n-2)S^2}{\sigma^2} \sim \chi_{n-2}^2$$

- S^2 er uavhengig av $\hat{\alpha}$

- ★ Husk definisjonen av t-fordeling: La $Z \sim N(0, 1)$ og $V \sim \chi_{\nu}^2$ uavhengige. Da er

$$T = \frac{Z}{\sqrt{\frac{V}{\nu}}} \sim t_{\nu}$$

- ★ Fra dette følger det at

$$\frac{\hat{\alpha} - \alpha}{\sqrt{\sigma^2 \cdot \frac{\sum_{i=1}^n x_i^2}{n \sum_{i=1}^n (x_i - \bar{x})^2}}} \sim N(0, 1) \quad \text{og} \quad \frac{\hat{\alpha} - \alpha}{\sqrt{S^2 \cdot \frac{\sum_{i=1}^n x_i^2}{n \sum_{i=1}^n (x_i - \bar{x})^2}}} = \frac{\frac{\hat{\alpha} - \alpha}{\sqrt{\sigma^2 \cdot \frac{\sum_{i=1}^n x_i^2}{n \sum_{i=1}^n (x_i - \bar{x})^2}}}}{\sqrt{\frac{(n-2)S^2}{\sigma^2}}}} \sim t_{n-2}$$

- ★ Dermed

$$\text{Pivotal: } \frac{\hat{\alpha} - \alpha}{\sqrt{S^2 \cdot \frac{\sum_{i=1}^n x_i^2}{n \sum_{i=1}^n (x_i - \bar{x})^2}}} \sim t_{n-2}, \quad \text{Testobservator: } T = \frac{\hat{\alpha} - \alpha_0}{\sqrt{S^2 \cdot \frac{\sum_{i=1}^n x_i^2}{n \sum_{i=1}^n (x_i - \bar{x})^2}}} \sim t_{n-2} \text{ når } H_0 \text{ er sann}$$

Inferens om stigningstallet β

- ★ Har tidligere funnet at

$$\hat{\beta} \sim N\left(\beta, \frac{\sigma^2}{\sum_{i=1}^n (x_i - \bar{x})^2}\right), \quad \frac{(n-2)S^2}{\sigma^2} \sim \chi_{n-2}^2$$

- S^2 er uavhengig av $\hat{\beta}$

- ★ Husk definisjonen av t-fordeling: La $Z \sim N(0, 1)$ og $V \sim \chi_{\nu}^2$ uavhengige. Da er

$$T = \frac{Z}{\sqrt{\frac{V}{\nu}}} \sim t_{\nu}$$

- ★ Fra dette følger det at

$$\frac{\hat{\beta} - \beta}{\sqrt{\frac{\sigma^2}{\sum_{i=1}^n (x_i - \bar{x})^2}}} \sim N(0, 1) \quad \text{og} \quad \frac{\hat{\beta} - \beta}{\sqrt{\frac{S^2}{\sum_{i=1}^n (x_i - \bar{x})^2}}} = \frac{\frac{\hat{\beta} - \beta}{\sqrt{\frac{\sigma^2}{\sum_{i=1}^n (x_i - \bar{x})^2}}}}{\sqrt{\frac{\frac{(n-2)S^2}{\sigma^2}}{n-2}}} \sim t_{n-2}$$

- ★ Dermed

$$\text{Pivotal: } \frac{\hat{\beta} - \beta}{\sqrt{\frac{S^2}{\sum_{i=1}^n (x_i - \bar{x})^2}}} \sim t_{n-2}, \quad \text{Testobservator: } T = \frac{\hat{\beta} - \beta_0}{\sqrt{\frac{S^2}{\sum_{i=1}^n (x_i - \bar{x})^2}}} \sim t_{n-2} \text{ når } H_0 \text{ er sann}$$

Inferens om $\mu_0 = E[Y|x_0] = \alpha + \beta x_0$

- ★ Estimator for μ_0 :

$$\begin{aligned}\hat{\mu}_0 &= \hat{\alpha} + \hat{\beta}x_0 \\ &= (\bar{Y} - \hat{\beta}\bar{x}) + \hat{\beta}x_0 \\ &= \bar{Y} + \hat{\beta}(x_0 - \bar{x})\end{aligned}$$

- ★ Egenskaper til $\hat{\mu}_0$

$$E[\hat{\mu}_0] = E[\hat{\alpha} + \hat{\beta}x_0] = E[\hat{\alpha}] + E[\hat{\beta}x_0] = E[\hat{\alpha}] + E[\hat{\beta}]x_0 = \alpha + \beta x_0 = \mu_0$$

$$\text{Var}[\hat{\mu}_0] = \text{Var}[\bar{Y} + \hat{\beta}(x_0 - \bar{x})] = \text{Var}[\bar{Y}] + \text{Var}[\hat{\beta}(x_0 - \bar{x})] + 2\text{Cov}[\bar{Y}, \hat{\beta}(x_0 - \bar{x})]$$

$$= \text{Var}\left[\frac{1}{n} \sum_{i=1}^n Y_i\right] + \text{Var}[\hat{\beta}](x_0 - \bar{x})^2 + 2(x_0 - \bar{x})\text{Cov}[\bar{Y}, \hat{\beta}]$$

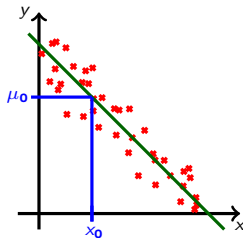
$$= \left(\frac{1}{n}\right)^2 \text{Var}\left[\sum_{i=1}^n Y_i\right] + \frac{\sigma^2}{\sum_{i=1}^n (x_i - \bar{x})^2} (x_0 - \bar{x})^2$$

$$= \frac{1}{n^2} \sum_{i=1}^n \text{Var}[Y_i] + \frac{\sigma^2 (x_0 - \bar{x})^2}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

$$= \frac{1}{n^2} \sum_{i=1}^n \sigma^2 + \frac{\sigma^2 (x_0 - \bar{x})^2}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

$$= \frac{\sigma^2}{n} + \frac{\sigma^2 (x_0 - \bar{x})^2}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

$$= \sigma^2 \left(\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \right)$$



- ★ $\hat{\mu}_0$ er en lineær funksjon av $Y_1, Y_2, \dots, Y_n$
- ★ $Y_1, Y_2, \dots, Y_n$ er uavhengige, normalfordelte
- ★ Dette gir at

$$\hat{\mu}_0 \sim N\left(\mu_0, \sigma^2 \left(\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum_{i=1}^n (x_i - \bar{x})^2}\right)\right)$$

- ★ Siden S^2 er uavhengig av $\hat{\alpha}$ og $\hat{\beta}$:
 - S^2 er uavhengig av $\hat{\mu}_0$

Pivotal for å utlede konfidensintervall for μ_0

- ★ Har tidligere funnet at

$$\hat{\mu}_0 \sim N\left(\mu_0, \sigma^2 \left(\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum_{i=1}^n (x_i - \bar{x})^2}\right)\right), \quad \frac{(n-2)S^2}{\sigma^2} \sim \chi_{n-2}^2$$

- S^2 er uavhengig av $\hat{\mu}_0$

- ★ Husk definisjonen av t-fordeling: La $Z \sim N(0, 1)$ og $V \sim \chi_\nu^2$ uavhengige. Da er

$$T = \frac{Z}{\sqrt{\frac{V}{\nu}}} \sim t_\nu$$

- ★ Fra dette følger det at

$$\frac{\hat{\mu}_0 - \mu_0}{\sqrt{\sigma^2 \left(\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum_{i=1}^n (x_i - \bar{x})^2}\right)}} \sim N(0, 1) \quad \text{og} \quad \frac{\hat{\mu}_0 - \mu_0}{\sqrt{S^2 \left(\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum_{i=1}^n (x_i - \bar{x})^2}\right)}} = \frac{\frac{\hat{\mu}_0 - \mu_0}{\sqrt{\sigma^2 \left(\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum_{i=1}^n (x_i - \bar{x})^2}\right)}}}{\sqrt{\frac{(n-2)S^2}{\sigma^2}}} \sim t_{n-2}$$

- ★ Dermed kan vi som pivotal bruke

$$\frac{\hat{\mu}_0 - \mu_0}{\sqrt{S^2 \left(\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum_{i=1}^n (x_i - \bar{x})^2}\right)}} \sim t_{n-2}$$

Oppsummering

- ★ Har funnet pivotal og testobservator som kan brukes for enkel lineær regresjon
 - for konstantledd α
 - for stigningstall β
- ★ Har funnet pivotal som kan brukes for enkel lineær regresjon
 - for $\mu_{\mathbf{0}} = E[Y|x_{\mathbf{0}}] = \alpha + \beta x_{\mathbf{0}}$