TMA4268 24.01.2019

M3 L2: linear regression

Model:

$$\beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix}$$

$$Y_i = x_i^T \beta + \varepsilon_i = 1 \cdot \beta_0 + x_{i1}\beta_1 + x_{i2}\beta_2 + \dots + x_{ip}\beta_p + \varepsilon_i$$

$$x_i^T = [\,1 \quad x_{i1} \quad x_{i2} \quad \cdots \quad x_{ip}\,]$$

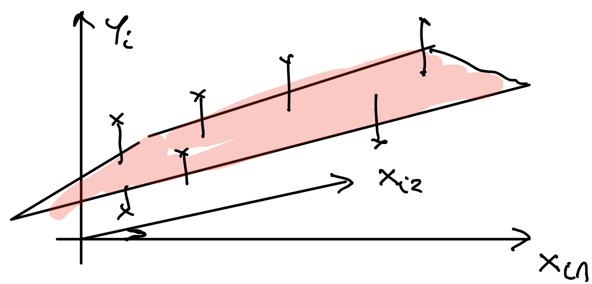
as row nr i of

$$\underbrace{Y}_{n\times 1} = \underbrace{X}_{n\times(p+1)}\underbrace{\beta}_{(p+1)\times 1} + \underbrace{\varepsilon}_{(n\times 1)}$$

Where $\varepsilon \sim N_n(0, \sigma^2 I)$

$E(\varepsilon) = 0$

$\underset{n\times n}{Cov(\varepsilon)} = \sigma^2 I$



This means that we assume that we have independent observation pairs (x_i, Y_i), $i = 1, .., n$.

PARAMETER ESTIMATION

$$\hat{\beta} = \underbrace{\underbrace{(\underbrace{X^T}_{(p+1)\times n}\underbrace{X}_{n\times(p+1)})^{-1}}_{(p+1)\times(p+1)}\underbrace{X^T}_{(p+1)\times n}}_{(p+1)\times n}\underbrace{Y}_{n\times 1}$$

$(p+1)$ linear combination of the Y_i

Example:

$\hat{\beta}_{year} = 0.026$

$\hat{\beta}_{kitch} = 1.14$

$X_{i,kitch} = \begin{cases} 0 & \text{standard} \\ 1 & \text{premium} \end{cases}$ dummy variable coding

If we compare two appartments where the only difference is that app. 1 has standard kitchen is from year a and app 2 has premium kitchen is from year a+1, then on average we expect that the rent sqm is 0.026 Euros 1.14 Euros higher for app 2 than app 1.

Distribution of $\hat{\beta} = (X^TX)^{-1}X^TY$

Formulas from M2 Part B: $Z = CY$, $E(Z) = C\,E(Y)$

$$Cov(Z) = C\,Cov(Y)\,C^T$$

Starting point: $Y = X\beta + \varepsilon \sim N_n(X\beta, \sigma^2 I)$

$\hat{\beta}$ will be multivariate normal with

$$E(\hat{\beta}) = E\left((X^TX)^{-1}X^TY\right) = (X^TX)^{-1}X^T\,\overbrace{E(Y)}^{X\beta}$$

$$= (X^TX)^{-1}X^TX\beta = \beta \quad \text{unbiased}$$

$$Cov(\hat{\beta}) = Cov\left((X^TX)^{-1}X^TY\right) = (X^TX)^{-1}X^T\,\overbrace{Cov(Y)}^{\sigma^2 I}\,\left[(X^TX)^{-1}X^T\right]^T$$

$$= \underbrace{(X^TX)^{-1}X^TX}\left((X^TX)^{-1}\right)^T\sigma^2 = \underline{\underline{(X^TX)^{-1}\sigma^2}}$$

NB: $(X^TX)^T = X^TX$ = symmetric

$(X^TX)^{-1}$ is also symmetric

So, $Var(\hat{\beta}_j) = \left[(X^TX)^{-1}\right]_{[j,j]}\sigma^2 = c_{jj}\,\sigma^2$

$Cov(\hat{\beta}_j, \hat{\beta}_k) = \left[(X^TX)^{-1}\right]_{[j,k]}\sigma^2$

Estimator for σ^2

$$Var(\varepsilon) = E(\varepsilon^2) - \underbrace{E(\varepsilon)}_{0}{}^2$$

residuals: $e_i = Y_i - \hat{Y}_i$

$= Y_i - x_i^T\hat{\beta}$

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{n} e_i^2}{n-p-1} = \frac{RSS}{n-\underbrace{p-1}_{\text{\# of regr. par. estimated!}}}$$

$$\frac{\hat{\sigma}^2(n-p-1)}{\sigma^2} \sim \chi^2_{n-p-1}$$

Ozone:

$$\mathrm{Cov}(\hat{\beta}_{temp}, \hat{\beta}_{wind}) = 0.08$$

<u>Inference</u> $\Leftarrow$ as for simple linear regression $\qquad \hat{\sigma}^2 = \frac{RSS}{n-p-1}$

"all" is based on $\qquad \hat{\beta}_j \sim N(\beta_j, \underbrace{(X^T X)^{-1}_{[j,j]}}_{c_{jj}} \sigma^2)$

$$T_j = \frac{\hat{\beta}_j - \beta_j}{\sqrt{c_{jj}}\,\hat{\sigma}} \sim t_{n-p-1}$$

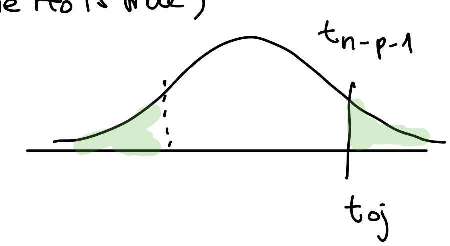
1) $(1-\alpha)\cdot 100\%$ CI : $\left[\hat{\beta}_j \pm t_{\frac{\alpha}{2}, n-p-1} \cdot \sqrt{c_{jj}} \cdot \hat{\sigma}\right]$

"never only give $\hat{\beta}_j$ - need also CI for β_j"

2) Single hypothesis testing: $H_0: \beta_j = 0$ vs $H_1: \beta_j \neq 0$

P-value $= 2 \cdot P(T_{0j} > |t_{0j}|$ given the H_0 is true$)$

$$T_{0j} = \frac{\hat{\beta}_j - 0}{\sqrt{c_{jj}} \cdot \hat{\sigma}} \qquad |t_{0j}| \;\|\; \text{numerical value in data}$$



<u>Is the regression significant?</u>

$H_0: \beta_1 = \beta_2 = \dots = \beta_p = 0$ vs H_1: at least one $\beta_j \neq 0$

(NB not β_0)

$\Rightarrow$ F-test

Prediction:

$x_0^T = [\,1 \quad x_{01} \quad x_{02} \ \cdots \ x_{0p}\,]$ new obs

prediction interval

$\hat{Y}_0 = x_0^T \hat{\beta}$ prediction, with PI

$(1-\alpha)\,100\%$ PI

$$Y_0 \in \left[\, x_0^T \hat{\beta} \pm t_{\frac{\alpha}{2}, n-p-1}\, \hat{\sigma} \sqrt{1 + x_0^T (\mathbb{X}^T \mathbb{X})^{-1} x_0} \,\right]$$