Kap 12: Den Simple Lineære Regressionsmodel: Modelkontrol.

25. maj 2002

Se evt. regneskema s. 378 for SAK det-ene-og–det-andet

Nr.	Hypotese	(A) Grafisk	(B) Numerisk	(C) Uddybende
-	**Modellen**	$\mathbf{y_i = \beta_0 + \beta_1 x_i + e_i}$ Opstil evt. m. opgaverelevante termer , eller omdøb variable, eks.: løn=x_1 og profit = y		
-	**Std.forudsætninger**	**I)** $E(e_i) = 0$ **II)** $Var(e_i) = \sigma^2$ for i =1,...,n. **III)** $Cov(e_i, e_j) = 0$ for $i \neq j \Rightarrow$ Uafhængige restled $\Rightarrow e_i \sim Nf(0,\sigma^2)$ …**IV)** hvid støj		
-	**Standardiserede residualer**	$e_i^* = \dfrac{e_i}{\sqrt{Var(e_i)}}$, hvor var (e_i) afhænger af værdien af den forklarende variabel Mens var(e_i*) er konstant (homoskedasticitet) for alle i = 1,...,n, så de kan sammenlignes direkte!		
1	Ad **I)** Linearitet	-Plot (x,y)-dia.: Kommenter lineær sammenhæng, - Plot (x,e_i*)-dia: Skal være hvid støj / usystematisk - Plot $(y,\hat{y})$ -dia.: skal ligge om 45º linie.	$R^2 = \dfrac{SAK_Y - SRK}{SAK_Y} > 0{,}5 \Rightarrow$ relativt høj	R^2= Determinationskoefficenten, angiver *den andel af den totale variation i responsvariablen (y), der kan beskrives udfra den lineære relation til den forklarende variabel (x)*
2	Ad. **II)** Homoskedasticitet	- Plot (x_i,e_i*): skal ligge usystematisk i et bredt bånd, med ca. α pct. uden for (x,|2,0|)	NIL	NIL
3	Ad. **III)** Uafhængige restled (Autokorr.-kontrol) *NB: Kun ved tidsrække obs.*	- Plot (t,e_i*): Kommenter ”*passede antal fortegnsskift*“	$DW = \dfrac{\sum_{i=2}^{n}(\hat{e}_i - \hat{e}_{i-1})^2}{\sum_{i=1}^{n}\hat{e}_i^{\,2}}$ 0 d_1 d_2 d_3 d_4 4 PA | ID | IA | ID | NA PA/NA=Pos./Neg. autokorr. ID=Indecisionsområdet IA = Ingen autokorrelation Forklar intuitivt hvorfor PA el. NA (hvis intuitivt)	- (t, ,e_i*)= bløde kurver ⇒ pos. autokorr. - (t, ,e_i*) =ZIG-ZAG ⇒ neg. autokorr. - DW kontrollerer for AR(1)-korrelation: dvs., om $e_i = \rho e_{i-1} + v_i$, *hvor* v_i er uaf. NF restled $\sim$ NF(0,σ^2)
4	Ad . **IV)** Normalfordelte std. residualer	Plot fraktil-dia. (Prob Plot): • Skal ligge *tilfældigt* om en ret linie	NIL	NIL
5	Betydning af enkelt-obs.	- Plot (x,y)-dia.: • Kommenter om nogle obs. ligger ”isoleret” og dermed påvirker estimationen meget • Kommenter hvorledes de påvirker parameter estimaterne.	Cook's afstand: $D_i = \dfrac{(\hat{\beta}-\hat{\beta}_{(i)})^T (X^T X)(\hat{\beta}-\hat{\beta}_{(i)})}{(p+1)s^2}$, $\hat{\beta}_{(i)}$ er MK parameterestimatet uden obs. nr. i.	- D_i = forskellen i parameterestimat med og uden obs. nr. i. - Anvendes til at udpege obs., der bør underkastes nærmere undersøgelse 1. Udpeg størst Cook’s D. 2. Sig hvorfor: Stor residual, ekstrem x-værdi eller begge. (evt. 3. Mahalanobis afstand: ydergrænse af obs.)
6	Test H_0:β_0= 0 vs. tosiddet…		$T = \hat{\beta}_0 / \sqrt{\text{var}(\hat{\beta}_0)} \sim t(n-2)$	
7	Test H_0:$\beta_1 = \beta_{10}$ vs. en-/tosiddet…		$T = (\hat{\beta}_1 - \beta_{10}) / \sqrt{\hat{\text{var}}(\hat{\beta}_1)} \sim t(n-2), \hat{\text{var}}(\hat{\beta}_1) = s^2 / SAK_x$	**Konf$_{1-\alpha}$[E(Y|x_0)]** = $\hat{y} \pm t_{1-\alpha/2}(n-2) \cdot s \cdot \sqrt{1/n + \text{var}(\hat{y})}$
8	H_0:R^2=0 vs. H_1: $R_2 \neq 0$ (kun tosiddet)		V=$R^2 \cdot$(n-2)/(1- R^2)~$F_{1-\alpha/2}$(1,n-2)	**Pred$_{1-\alpha}$(Y|x_0)** = $\hat{y} \pm t_{1-\alpha/2}(n-2) \cdot s \cdot \sqrt{1 + 1/n + \text{var}(\hat{y})}$
…	Mahalonobis afstand	Angiver grænser for intra- og ekstrapolation. Stor Mahalanobis afstand, men lille residual kan være med til at sikre modellens persistens/styrke (M_i mest anvendt i multiple regr. hvor man ikke bare kan se på værdien, men ekstrapolationsgrænsen bliver en ellipsoide)		