to je další ukázka <u>fit / prediction dilema</u>

- stejný efekt nastává také pro

$$\hat{\beta}_i - \hat{\beta}_{(-i)} = (X^T X)^{-1} x_i \hat{e}_{(-i)}$$

rozdíl může být „malý" pokud je „fit" dobrý, ale může být „velký" pokud je h_{ii} velké.

Míry influence

- i pro perfektní model mohou dva různé vzorky (X, y) a (X', y') vést k různým závěrům
- většinou máme k dispozici jen originální data
- bude nás zajímat vliv i-tého řádku X na model
- už víme: velké h_{ii} indikuje, že i-té poz. má velký vliv a velká rezidua naznačují možnou neadekvátnost modelu
- míry, které zavedeme, budou kombinovat tyto dva faktory
- podobné přístupy z PRESS reziduí, tam budeme sledovat jak velký vliv má vynechání i-tého pozorování na $\hat{\beta}$ a $\hat{y}_i$

DFBETAS

$\hat{\beta} - \hat{\beta}_{(-i)}$ měří vliv vynechání i-tého poz. na odhad β
(bude základem pro naši analýzu)

příp. $\hat{\beta} - \hat{\beta}_{(-i)} = (X^T X)^{-1} x_i \dfrac{\hat{e}_i}{1 - h_{ii}}$

a) vliv i-tého pozorování na β_j

$$\hat{\beta}_j - \hat{\beta}_{(-i)j} = \frac{r_{ji} \hat{e}_i}{1 - h_{ii}}$$

, kde r_{ji} je (j,i)-tý prvek matice $R = (X^T X)^{-1} X^T$

i - té pozorování považovat za influenční, na β_j pokud

$\hat\beta_j - \hat\beta_{(-i)j}$ bude velké

protože $\hat\beta_j$ je n. rozd., „velké" bychom měli měřit relativně

vzhledem k s.d. $(\hat\beta_j)$ což je $\sigma\sqrt{v_j}$, $v_j = (X^TX)^{-1}_{jj}$

pokud je odhadnuto pomocí $\hat\sigma_{(-i)}\sqrt{v_j}$, dostaneme definici

$$DFBETAS_{ji} = \frac{\hat\beta_j - \hat\beta_{(-i)j}}{\hat\sigma_{(-i)}\sqrt{v_j}} = \frac{v_j\, i\, \hat e_i}{\sqrt{v_j}\,\hat\sigma_{(-i)}(1-h_{ii})} = \frac{v_{ji}}{\sqrt{v_j}}\,\frac{\hat t_i}{\sqrt{1-h_{ii}}}$$

kde $\hat t_i$ je ext. studentizované residuum

- kombinuje efekt velkého residua $\hat t_i$ a velkého h_{ii}
- jedna možnost pro limitní hodnotu:

i - té pozorování je považováno za influenční na odhad β_j,

pokud $\quad |DFBETAS_{ji}| > \frac{2}{\sqrt{n}}$

- máme $(n+1)\times n$ hodnot pro rozhodnutí, zjednodušíme

b) vliv i - tého pozorování na celý vektor $\hat\beta$

použít nějaké normy na vektor $\hat\beta - \hat\beta_{(-i)}$

Cook navrhl $\quad D_i = \dfrac{(\hat\beta - \hat\beta_{(-i)})^T M (\hat\beta - \hat\beta_{(-i)})}{(m+1)c}$, kde M je PD matice a c je nějaká konst.

nejčastěji volba je $M = X^TX$ a $c = s_m^2$

Cookova vzdálenost:

$$D_i = \frac{(\hat\beta - \hat\beta_{(-i)})^T X^TX (\hat\beta - \hat\beta_{(-i)})}{(m+1)\, s_m^2}$$

$$D_i = \frac{1}{(m+1)\hat{\sigma}_m^2}\left(\frac{\hat{e}_i}{1-h_{ii}}\right)^2 \underbrace{x_i^T (X^TX)^{-1} X^T X (X^TX)^{-1} x_i}_{I} = \frac{1}{m+1}\frac{h_{ii}}{1-h_{ii}}\underbrace{\frac{\hat{e}_i^2}{\hat{\sigma}_m^2(1-h_{ii})}}_{(\hat{r}_i)^2}$$

$I = h_{ii}$

- vyjádření bez normy

$$D_i = \frac{\hat{r}_i^2}{m+1}\left(\frac{h_{ii}}{1-h_{ii}}\right)$$

Pozn: $100(1-\alpha)\%$ simultánní IS pro β je

$$C(\alpha) = \left\{\beta \;\middle|\; \frac{(\hat{\beta}-\beta)^T X^T X (\hat{\beta}-\beta)}{(m+1)\hat{\sigma}_m^2} \le F_{1-\alpha}(m+1, n-m-1)\right\}$$

Pozn. $\beta_{(-i)} \in C(\alpha) \Leftrightarrow D_i \le F_{1-\alpha}(m+1, n-m-1)$

Lze je možné pro RULE OF THUMB:

i-té pozorování je influenční, jestliže $D_i > F_{\frac{1}{2}}(m+1, n-m-1)$

(pro větší n, n $F_{\frac{1}{2}} \approx 1$, zjednodušeně pravidlo $D_i > 1$)

Pozn: Jako platí

$$D_i = \frac{(\hat{y} - \hat{y}_{(-i)})^T (\hat{y} - \hat{y}_{(-i)})}{(m+1)\hat{\sigma}_m^2}$$

Pozn. dá se chápat jako míra influence na celkovou ~~predikci~~ chybu

DFFITS: míra i-tého pozorování na $\hat{y}_i$

$$\text{DFFITS}_i = \frac{\hat{y}_i - \hat{y}_{(-i)}}{\hat{\sigma}_{(-i)}\sqrt{h_{ii}}} = \dots = \hat{r}_i\sqrt{\frac{h_{ii}}{1-h_{ii}}}$$

odhad s.d. $\hat{e}_i$

RULE OF THUMB: i-té pozorování je influenční, pokud $|\text{DFFITS}_i| > 3\sqrt{\frac{m+1}{n-m-1}}$

Pozn: Míry influence v R:

DFBETAS – dfbetas( ) DFFITS – dffits( )

Cookova vzdálenost D_i – cooks.distance( )

Leverage h_{ii} – hatvalues( )

vše shrnuje funkce influence.measures( )

(má navíc (covariance ratio))

používané pravidla:

i-té pozor. je influenční pokud:

$$|DFBETAS| > 1, \quad |DFFITS| > 3\sqrt{\frac{m+1}{n-m-1}},$$

$$D_i > F_{0,5}(m+1, n-m-1), \quad h_{ii} > 3\frac{m+1}{n}$$

TRANSFORMACE

- pokud není splněn některý z předpokladů modelu (linearita, normalita chyb, homoskedasticita) jednou z možností je pokusit se transformovat nějaké proměnné, aby transformovaný model tyto předpoklady alespoň „přibližně" splňoval

Transformace vysvětlované proměnné y

Hledáme funkci $h(\cdot)$ tak, aby model $Y_i^* = h(Y_i) = \beta_0 + \sum_{j=1}^{m} x_{ij}\beta_j + \varepsilon_i$ splňoval předpoklady

3 hlavní důvody pro transformaci Y:

1) transformace účelu měření tak, aby odpovídala celé R, což může odstranit problémy s podmínkami na β

např. studie kapacity plic (FEV dolní, FEV > 0)

- chtěli bychom, aby model nepredikoval záporné hodnoty
  (⇒ restriktivní podmínky na par. β)

  lze obejít modelováním $y^* = \log FEV$

pokud y jsou počty a 0 je možná hodnota, často se používá

$$y^* = \log(y+1) \quad \text{nebo obecně} \quad y^* = \log(y+c)$$

2) transformace Y, aby její rozdělení bylo „více" normální

- typicky to znamená pokusit se udělat rozdělení hodnot y více symetrické

často se setkáváme s rozděleními vychýlenými vpravo
(obvykle se to stává, pokud měříme nějakou fyz. veličinu, která může nabývat pouze kladných hodnot)

transformace $y^* = \log y$ nebo $y^* = y^\lambda, \lambda < 1$ budou redukovat toto vychýlení

typický postup: začít s hodnotou λ blízkou 1, pak snižovat hodnotu λ dokud není dosaženo „přibližně" symetrie reziduí

3) možná nejzásadnější motivace je pokusit se dosáhnout konstantní rozptyl přes všechna pozorování

- např. pro fyzikální veličiny s kladnými hodnotami se často stane, že rozptyl bude malý pro $\mu \approx 0$ a větší pro μ velké
  (už jen z důvodu, že obor hodnot y je omezen na kladné hodnoty)

říká se tomu *positive mean-variance relationship*

- nepřesnost měření kladných veličin se také často vyjadřuje

pomocí koeficientu variace $CV(Y) = \frac{\text{s.d. } Y}{E(Y)}$

čísla bývají více konstantní mezi různými než s.d.

veličinách vyjadřujeme relativní spíše než absolutní

motivováni to znamená, že $\underline{\text{Var } Y = \phi\, E(Y)^2 = \phi\, \mu^2}$ pro nějaké ϕ

– pro odstranění vztahu mezi $E(Y)$ a $\text{Var } Y$ se často používají

mocninné transformace $\eta^* = \eta^2$ (pro $\eta > 0$)

Transformace:	← …	η^3	η^2	η	$\sqrt{\eta}$	$\log \eta$	$\frac{1}{\sqrt{\eta}}$	$\frac{1}{\eta}$	$\frac{1}{\eta^2}$ … →
Box-Cox λ:		3	2	1	$\frac{1}{2}$	0	$-\frac{1}{2}$	-1	-2

- pokud $\text{Var } Y$ klesá s rostoucí EY
- pokud $\text{Var } Y$ roste s rostoucí EY

OBECNĚ: předpokládejme vztah $\underline{\text{Var } Y = \phi\, V(\mu)}$

a uvažujme transformaci $\eta^* = h(\eta)$

Taylorův rozvoj 1. řádu funkce $h(\eta)$ v bodě μ

$$\eta^* = h(\eta) \approx h(\mu) + h'(\mu)(\eta - \mu)$$

z čehož plyne, že $\text{Var } Y^* \approx (h'(\mu))^2 \cdot \text{Var } Y$

transformace $\eta^* = h(\eta)$ bude bude přibližně stabilizovat rozptyl,

pokud $\underline{h'(\mu)}$ je ~~přímo~~ úměrné $\underline{(\text{Var } Y)^{-\frac{1}{2}} = V(\mu)^{-\frac{1}{2}}}$

- pokud $V(\mu) = \mu^2$ $\Rightarrow$ transformace stabilizující rozptyl je $\log(\eta) = h(\eta)$, protože $h'(\mu) = \frac{1}{\mu}$
- pokud $V(\mu) = \mu$ $\Rightarrow$ stab. transformace je $h(\eta) = \sqrt{\eta}$, protože $h'(\mu) = \frac{1}{2\sqrt{\mu}}$

$$\left(h(\mu) = \int \frac{d\mu}{\sqrt{V(\mu)}} \right)$$

- asi nejvíce užívaná transformace je $y^* = \log(y)$
  jedním z důvodů je i dobrá interpretovatelnost par. β

Pozn (Interpretace parametrů LM):

a) klasický LM: $E y = \beta_0 + \beta_1 x_1 + \dots + \beta_m x_m$

jednotková změna proměnné x_j $\Rightarrow$ změna Ey o β_j jednotek
(při ostatních proměnných stejných)

$x = (1, x_1, \dots, x_m)$, $x_{new} = (1, x_1, \dots, x_j+1, \dots x_m)$

$\downarrow$ Ey $\qquad$ $\downarrow$ Ey_{new}

$E y_{new} - Ey = \beta_j$

b) LM pro $\log y$: $\log y = \beta_0 + \beta_1 x_1 + \dots + \beta_m x_m + e$

$e \sim N(0, \sigma^2 I)$

pokud je to správný model, znamená to, že

$\log y \sim N(\mu, \sigma^2 I) \Rightarrow y \sim LN(\mu, \sigma^2)$ a platí

$Ey = e^{\mu + \frac{\sigma^2}{2}}$

predikce pro $E \log y$ je $\hat{\mu} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \dots + \hat{\beta}_m x_m$

predikce pro Ey bude $e^{\hat{\beta}_0 + \hat{\beta}_1 x_1 + \dots + \hat{\beta}_m x_m + \frac{\hat{\sigma}^2}{2}}$

uvažujme opět jednotkovou změnu p. x_j $(x_j \to x_j + 1)$

$$\frac{E y_{new}}{Ey} = \frac{e^{\hat{\beta}_0 + \hat{\beta}_1 x_1 + \dots + \hat{\beta}_j x_j + \hat{\beta}_j + \dots + \hat{\beta}_m x_m + \frac{\hat{\sigma}^2}{2}}}{e^{\hat{\beta}_0 + \hat{\beta}_1 x_1 + \dots + \hat{\beta}_m x_m + \frac{\hat{\sigma}^2}{2}}} = e^{\hat{\beta}_j}$$

bez závislosti

jednotková změna proměnné x_j $\Rightarrow$ multiplikativní změna Ey $e^{\hat{\beta}_j}$-krát

jiná možnost: $100(e^{\hat{\beta}_j} - 1)$ je procentní změna Ey způsobená jednotkovou změnou x_j