Rezidua, diagnostika a transformace

- je třeba ověřit adekvátnost modelu
  máme R^2, t, F statistiky, ty ale byly odvozeny za předpokladu
  linearity modelu a dalších podmínek na náhodné chyby
  pro ověření – analýza reziduí
  je důležitým nástrojem
- je také třeba ověřit vliv jednotlivých pozorování na model
  analýza odlehlých (outliers) a influenčních pozorování

  (velké reziduum pro i-té poz. ~~sch~~ naznačuje problém s modelem
  ale může to být i naopak, vlivné poz. nemusí mít velké reziduum)

  obr. 3.9 str. 111
- pokud detekujeme nějaké problémy s modelem, mohou pomoci
  transformace proměnných nebo metody na korekci nekonstantního rozptylu

Rezidua

připomenutí: $\hat{y} = X\hat{\beta} = Hy$, kde $H = X(X^TX)^{-1}X^T$

$$\hat{e} = y - \hat{y} = (I_n - H)y = (I_n - H)e$$

dále jsme ukázali: $E\hat{e} = 0$, $Cov(\hat{e}) = \sigma^2(I_n - H)$

pokud navíc $e \sim N_n(0, \sigma^2 I_n)$, potom $\hat{e} \sim N_n(0, \sigma^2(I_n - H))$

pokud označíme $h_{ii} = H_{ii}$, $\hat{e}_i \sim N(0, \sigma^2(1-h_{ii}))$, $Cov(\hat{e}_i, \hat{e}_j) = -\sigma^2 h_{ij}$

obecně bývá vhodnější pracovat se standardizovanými rezidui,

protože $Var(\hat{e}_i) = \sigma^2(1-h_{ii})$, pro $r_i = \dfrac{\hat{e}_i}{\sigma\sqrt{1-h_{ii}}}$ platí $Var(r_i) = 1$

σ odhadneme pomocí $s_n = \sqrt{\dfrac{1}{n-m-1} SSE}$, dostaneme

$$\hat{r}_i = \frac{\hat{e}_i}{s_n \sqrt{1-h_{ii}}}, \quad i = 1, \dots, n$$

interně studentizovaná rezidua

(někdy se nazývají standardizovaná rezidua)

R: rstandard(.)

pokud σ^2 odhadneme na základě modelu, ve kterém bylo vynecháno i-té pozorování, označíme tento odhad $\hat{\sigma}^2_{(-i)}$, potom

$$\hat{t}_i = \frac{\hat{e}_i}{\hat{\sigma}_{(-i)} \sqrt{1-h_{ii}}}, \quad i = 1, \dots, n$$

externě studentizovaná rezidua

(někdy se nazývají studentizovaná)

R: rstudent(.)

např. $\hat{\sigma}^2_{(-i)} = \frac{SSE_{(-i)}}{n-p-2}$ je nestranný odhad σ^2 v modelu $(-i)$

POZN:
- pokud h_{ii} je malé, pro velké n by se měla $\hat{e}_i, \hat{r}_i, \hat{t}_i$ chovat přibližně stejně a $\hat{r}_i, \hat{t}_i \approx N(0,1)$
- pro malá n ($n < 20$) a/nebo $h_{ii} \approx 1$ je preferováno použít $\hat{r}_i$ nebo $\hat{t}_i$ a aktuálně bývá častěji ~~používáno~~ doporučováno $\hat{t}_i$ (i-té poz. s velkým h_{ii} může rozjet odhad σ^2 a tím snižuje velikost svého rezidua)
- h_{ii} hrají zásadní roli v diagnostice modelu, podíváme se na jejich základní vlastnosti

Leverage h_{ii} - potenciál i-tého poz.

leverage point - pákový bod, vzdálený bod

- $\text{Var}\, \hat{e}_i = \sigma^2 (1 - h_{ii}) \geq 0 \Rightarrow h_{ii} \leq 1$

- $H^2 = H \Rightarrow h_{ii} = \sum_{j=1}^{n} h_{ij} h_{ji} = \sum_{j=1}^{n} (h_{ij})^2$, tedy $h_{ii} > 0$

(dá se ukázat silnější tvrzení: $h_{ii} \geq \frac{1}{n}$)

- $\underbrace{HX}_{A} = X \Rightarrow A_{i1} = \sum_{j=1}^{n} h_{ij} \underbrace{x_{j1}}_{=1} = \sum_{j=1}^{n} h_{ij} = x_{i1} = 1$, tedy

$$\sum_{j=1}^{n} h_{ij} = 1 \quad \forall i \in \hat{n}$$

- význam h_{ii} vyplývá z následující úvahy:

$$\hat{y} = H y \Rightarrow \hat{y}_i = \sum_{j=1}^{n} h_{ij} y_j = h_{ii} y_i + \sum_{\substack{i \neq j \\ j=1}}^{n} h_{ij} y_j$$

pokud $h_{ii} \approx 1$, potom $\hat{y}_i \approx y_i$ a model je nucen proložit přímku bodem (x_i, y_i) i když tam nepatří

body s „velkým h_{ii}" – body s velkým potenciálem (high leverage points)

Tyto body by měly být detekovány pro další zkoumání

~~pokud y_i je bod s velkým potenciálem, potom $\hat{e}_i$ může být malé, ale $\hat{h}_i$ je „velké", tedy potenciál poz. y_i maskuje efekt stálého modelu na y_i~~

- otázka je, jaká hodnota h_{ii} je „velká"

heuristické pravidlo:

$$\sum_{i=1}^{n} h_{ii} = \operatorname{tr} H = m+1, \quad \text{tzn. } \frac{m+1}{n} \text{ je průměrná hodnota } h_{ii}$$

i-té poz. má velký potenciál, jestliže $h_{ii} > \frac{3(m+1)}{n}$

(tohle dělá i R)

Grafy residuí

A) ověření normality – histogramy, Q-Q plots

tyto obrázky nezávisí na počtu nezávislých proměnných x,

vše stejné jako v jednoduché LR

B) pro ověření funkční formy pro $E(Y_x)$ a/nebo konstantního rozptylu se nejčastěji používají:

1) grafy $\hat{e}_i, \hat{r}_i$ nebo $\hat{t}_i$ oproti $\mathbb{x}_j^c$, $j = 1, \dots, m$

kde $\mathbb{x}_j^c$ je j-tý sloupec $\mathbb{X}$

2) grafy $\hat{e}_i, \hat{r}_i$ nebo $\hat{t}_i$ oproti $\hat{y}_i$

3) partial residual plots

POZN.: • zdůvodnění 1):

normální rovnice $\mathbb{X}^T(y - \mathbb{X}\hat{\beta}) = 0$ implikují $\mathbb{X}^T(y - \hat{y}) = \mathbb{X}^T\hat{e} = 0$

připomenutí: $Y_i = \beta_1 x_i + e_i$, $\hat{\beta}_1 = \dfrac{\sum_{i=1}^n x_i y_i}{\sum x_i^2} = \dfrac{\mathbb{x}^T y}{\|\mathbb{x}\|^2}$

pokud tedy uvažujeme LR model bez interceptu pro $\hat{e}$ v závislosti na $\mathbb{x}_j^c$, směrnice přímky bude

$$\hat{\beta}_j^* = \frac{(\mathbb{x}_j^c)^T \hat{e}}{\|\mathbb{x}_j^c\|^2} = 0$$

graf $\hat{e}_i, \hat{r}_i, \hat{t}_i$ proti $\mathbb{x}_j^c$ by měl dávat náhodně rozptýlené body kolem osy x (bez trendu, $\hat{r}_i, \hat{t}_i$ uvnitř přibl. ± 2)

pokud tomu tak není, může to signalizovat nelinearitu v $\mathbb{x}_j$ nebo nekonstantní rozptyl

- zdůvodnění 2):

ukázali jsme $\sum_{i=1}^{n} \hat{y}_i \hat{e}_i = 0$

pro LM bez interceptu pro $\hat{e}_i$ oproti $\hat{y}_i$ tedy platí

$$\hat{\beta} = \frac{\hat{e}^T \hat{y}}{\|\hat{y}\|^2} = 0$$

body by opět měly být náh. rozptýlené kolem osy x

- trychtýřovitý tvar indikuje nekonstantní rozptyl
- ~~tvar~~ trendy indikují nelinearitu

3) Partial residual plot

- i tyto grafy $\hat{e}_i$ oproti x_j^c a $\hat{y}$ mohou indikovat nedostatky modelu, nemusí ~~tak~~ být zřejmé, jaké ty nedostatky jsou
- v SLR graf $\hat{e}_i$ proti x_i lze použít pro detekci nelinearity
- ale v MLR tyto grafy, stejně jako scatterploty, mohou být zavádějící, protože $\hat{e}$ závisí na všech prediktorech, nemusí být tedy izolován efekt dané proměnné při odstranění efektu ostatních
- pro zkoumání efektu j-té proměnné lze použít partial residual plots
  - lze je chápat jako ekvivalent scatterplotu v SLR

def. $\hat{e}_j^* = \hat{e} + \hat{\beta}_j x_j^c$, kde $\hat{e}$ je vektor reziduí modelu
$\hat{\beta}_j$ je LSE par. β_j
x_j^c · j-tý sloupec X

PRP: graf $\hat{e}_j^*$ oproti x_j^c, $j = 1, \dots, m$

pokud je model správný, měly by být body náhodně rozmístěné kolem přímky se směrnicí $\hat{\beta}_j$

zdůvodnění:

vztah mezi $\hat{e}_j^*$ a x_j^c má formu SLR bez interceptu

pokud je model správný, $\hat{e}_i$, $i = 1, \dots, n$, splňují $E\hat{e}_i = 0$, $\mathrm{Var}\,\hat{e}_i = \sigma^2(1-h_{ii})$

má tedy smysl uvažovat RM pro $\hat{e}_j^*$ oproti x_j^c ($\hat{e}_j^* = \gamma_j x_j^c + e$)

pro odhad koef. platí

$$\underline{\hat{\gamma}_j} = \frac{(\hat{e}_j^*)^T x_j^c}{\|x_j^c\|^2} = \frac{(\hat{e} + \hat{\beta}_j x_j^c)^T x_j^c}{\|x_j^c\|^2} = \frac{\hat{e}^T x_j^c + \hat{\beta}_j \|x_j^c\|^2}{\|x_j^c\|^2} = \underline{\hat{\beta}_j}$$

protože $\hat{e}^T x_j^c = 0$

Př. housing price data str. 253

Př. birth weight data Ex 6.2, str. 255

POZN: PRPs jsou někdy kritizovány za nadhodnocování efektu x_j^c

alternativa: partial regression plot (added variable plot)

motivace: ptáme se, zda přidat novou proměnnou do modelu a chtěli bychom odhadnout její efekt

budeme tedy uvažovat rozšířený model

$$\underline{Y = X\beta + \gamma w + e} \qquad w \text{ je nový vektor regresorů}$$

lze zapsat jako

$$Y = [X\ w]\begin{pmatrix}\beta\\ \gamma\end{pmatrix} + e = X_w \beta_w + e$$

použitím normálních rovnic pro X_w lze odvodit formuli pro $\hat{\gamma}$

$$\underline{\hat{\gamma} = \frac{\hat{e}^T (I-H) w}{\|(I-H) w\|^2}} \quad (\#)$$

$\hat{\gamma}$ je směrnice RM pro $\hat{e}$ v závislosti na $w_{res} = (I-H)w$

(rezidua modelu pro w v závislosti na X)

teď uvažujme naopak, že w je sloupec původní X, řekněme x_j^c

a ozn. $X_{(-j)}$ matici X bez sloupce j

v předchozím modelu položme $\mathbb{X} = \mathbb{X}_{(-j)}$ a $w = x_j^c$

potom LSE $\hat{\beta}_j$ pro β_j je

$$\hat{\beta}_j = \frac{\hat{e}_{(-j)}^T \, x^c_{j,res}}{\| x^c_{j,res} \|^2}$$

, kde $\hat{e}_{(-j)}$ jsou rezidua modelu bez x_j^c

$x^c_{j,res} = (I - H_{(-j)}) \, x_j^c$, tedy jsou to rezidua modelu pro x_j^c v závislosti na ostatních proměnných, tedy na $\mathbb{X}_{(-j)}$

(v $x^c_{j,res}$ je tedy odstraněn efekt ostatních regresorů)

je to směrnice RM pro $\hat{e}_{(-j)}$ v závislosti na $x^c_{j,res}$

$\Downarrow$

added variable plot: graf $\hat{e}_{(-j)}$ proti $x^c_{j,res}$, $j = 1, \dots, m$

pokud je model správný, body by měly být náhodně rozptýlené kolem přímky se směrnicí $\hat{\beta}_j$ proch. počátkem

pokud závislost na x_j není lineární, projeví se to odklonem bodů od přímky

POZN: Ze vztahu (#) je vidět, že MLR může být chápána jako posloupnost SLR, kde postupně vkládáme modely pro novou proměnnou s použitím reziduí ~~pro~~ modelu pro předcházející proměnné

PRESS rezidua (PRESS residual, deleted residuals)

- pokud budeme chtít model použít nejen k vysvětlení vztahu mezi proměnnými ale také pro predikci, hodila by se míra vyjadřující jak dobře model predikuje (doposud jsme zkoumali jen jak dobře popisuje)
- šlo by použít IS nebo IP, to bychom ale předem museli znát body ve kterých chceme predikovat

- nejjednodušší přístup, jak měřit prediktivní přesnost modelu, by byla analýza reziduí pro predikce hodnot v nových bodech x, obecně ale nemáme data (y) v těchto bodech
- jedna možnost je použít data, která máme k dispozici

postup: vynecháme jedno pozorování, sestavíme model bez tohoto pozorování a porovnáme predikovanou a pozorovanou hodnotu pro vynechané pozorování

předp., že vynecháme i-té pozorování, ozn. $\hat{\beta}_{(-i)}$ odhad β v modelu s vynechaným i-tým pozorováním ($M_{(-i)}$) a $\hat{y}_{(-i)}$ predikovanou hodnotu modelem $M_{(-i)}$ v bodě x_i^T,

tzn. $$\hat{y}_{(-i)} = x_i^T \hat{\beta}_{(-i)}$$

potom $$\hat{e}_{(-i)} = y_i - \hat{y}_{(-i)}, \quad i = 1, \dots, n$$

nazýváme i-té PRESS reziduum

$$PRESS = \sum_{i=1}^{n} \hat{e}_{(-i)}^2$$ je užitečná míra přesnosti predikce

POZN: Otázka je jak počítat $\hat{e}_{(-i)}$, $i = 1, \dots, n$

- pro n velké se zdá, že to bude náročný problém, protože pro každé $i \in n$ musíme sestavit nový model
- naštěstí to není nutné, ukážeme totiž, že
$$\hat{e}_{(-i)} = \frac{\hat{e}_i}{1 - h_{ii}}$$

tzn. všechna $\hat{e}_{(-i)}$ lze snadno spočítat pomocí reziduí a hodnot h_{ii} z původního (úplného) modelu

Označme $\mathbf{x}_i^T$ - i-tý řádek matice $\mathbb{X}$

$\mathbb{X}_{(-i)}$ - matice $\mathbb{X}$ bez i-tého řádku

Věta: Jestliže $h_{ii} \neq 1$, potom

$$\left[\mathbb{X}_{(-i)}^T \mathbb{X}_{(-i)}\right]^{-1} = (\mathbb{X}^T\mathbb{X})^{-1} + \frac{(\mathbb{X}^T\mathbb{X})^{-1} \mathbf{x}_i \mathbf{x}_i^T (\mathbb{X}^T\mathbb{X})^{-1}}{1-h_{ii}},$$

kde h_{ii} je i-tý diag. prvek matice H.

Dk: nejdříve ukážeme

$$\underset{(m+1)\times(m+1)}{\mathbb{X}^T\mathbb{X}} = \underset{(m+1)\times(m+1)}{\mathbb{X}_{(-i)}^T \mathbb{X}_{(-i)}} + \underset{(m+1)\times(m+1)}{\mathbf{x}_i \mathbf{x}_i^T} \qquad (+)$$

kvůli značení předp. $i = n$, toho se dá vždy dosáhnout permutací řádků $\mathbb{X}$

potom

$$(\mathbb{X}^T\mathbb{X})_{ij} = \sum_{k=1}^{n} x_{ki}x_{kj} = \sum_{k=1}^{n-1} x_{ki}x_{kj} + x_{ni}x_{nj}$$

i,j-tý prvek $\mathbb{X}_{(-n)}^T \mathbb{X}_{(-n)}$ je $\sum_{k=1}^{n-1} x_{ki}x_{kj}$

i,j-tý prvek $\mathbf{x}_n \mathbf{x}_n^T$ je $x_{ni}x_{nj}$, tzn. (+) platí

Věta z LA: (Sherman - Morrison - Woodbury)

Nechť A je $n\times n$ invertibilní matice a nechť $\mathbf{u}$ je $n\times 1$ sloupcový vektor. Jestliže $\mathbf{u}^T A^{-1}\mathbf{u} \neq 1$, potom matice $B = A - \mathbf{u}\mathbf{u}^T$ je invertibilní a platí

$$B^{-1} = A^{-1} + \frac{A^{-1}\mathbf{u}\mathbf{u}^T A^{-1}}{1-\mathbf{u}^T A^{-1}\mathbf{u}}$$

položme $A = X^T X$, $u = x_i$, $B = X_{(-i)}^T X_{(-i)}$

tzn. $B = A - u u^T$, A je invertibilní a

$$u^T A^{-1} u = x_i^T (X^T X)^{-1} x_i = (X(X^T X)^{-1} X^T)_{ii} = h_{ii} \neq 1$$

užitím věty dostaneme

$$(X_{(-i)}^T X_{(-i)})^{-1} = (X^T X)^{-1} + \frac{(X^T X)^{-1} x_i x_i^T (X^T X)^{-1}}{1 - h_{ii}} \qquad \text{Q.E.D.}$$

Věta: nechť $\hat{e}_{(-i)}$ je i-té PRESS reziduum. Potom

$$\underline{\hat{e}_{(-i)} = \frac{\hat{e}_i}{1 - h_{ii}}}, \quad i = 1, \dots, n.$$

Dk. nechť $\hat{\beta}_{(-i)}$ je odhad β v modelu $M_{(-i)}$, tzn.

$$\hat{\beta}_{(-i)} = (X_{(-i)}^T X_{(-i)})^{-1} X_{(-i)}^T y_{(-i)}, \quad \text{kde } y_{(-i)} \text{ je } y \text{ bez } i\text{-té složky}$$

tzn.

$$\underline{\hat{y}_{(-i)}} = x_i^T \hat{\beta}_{(-i)} = x_i^T \underbrace{(X_{(-i)}^T X_{(-i)})^{-1}}_{(X^T X)^{-1} + \frac{(X^T X)^{-1} x_i x_i^T (X^T X)^{-1}}{1 - h_{ii}}} X_{(-i)}^T y_{(-i)} =$$

$$= x_i^T (X^T X)^{-1} X_{(-i)}^T y_{(-i)} + \frac{1}{1 - h_{ii}} x_i^T (X^T X)^{-1} x_i x_i^T (X^T X)^{-1} X_{(-i)}^T y_{(-i)} =$$

$$\underline{= S_1 + \frac{1}{1 - h_{ii}} S_2}$$

protože $X_{(-i)}^T y_{(-i)} = X^T y - y_i x_i$ dostaneme

$$\underline{S_1} = x_i^T (X^T X)^{-1} (X^T y - y_i x_i) = x_i^T \underbrace{(X^T X)^{-1} X^T y}_{\hat{\beta}} - y_i \underbrace{x_i^T (X^T X)^{-1} x_i}_{h_{ii}} =$$

$$= x_i^T \hat{\beta} - h_{ii} y_i = \underline{\hat{y}_i - h_{ii} y_i}$$

podobně

$$\underline{S_2} = \underbrace{x_i^T (X^TX)^{-1} x_i}_{h_{ii}} \underbrace{x_i^T (X^TX)^{-1} X^T y}_{\hat{y}_i} - y_i \underbrace{x_i^T (X^TX)^{-1} x_i}_{h_{ii}} \underbrace{x_i^T (X^TX)^{-1} x_i}_{h_{ii}} =$$

$$= \underline{h_{ii} \hat{y}_i - y_i h_{ii}^2}$$

tedy

$$\hat{y}_{(-i)} = \hat{y}_i - h_{ii} y_i + \frac{1}{1-h_{ii}} (h_{ii} \hat{y}_i - y_i h_{ii}^2)$$

celkem tedy

$$\hat{e}_{(-i)} = y_i - \hat{y}_{(-i)} = y_i (1+h_{ii}) - \hat{y}_i - \frac{1}{1-h_{ii}} (h_{ii} \hat{y}_i - y_i h_{ii}^2)$$

$$= \frac{1}{1-h_{ii}} \left(y_i (1-h_{ii}^2) - \hat{y}_i (1-h_{ii}) - h_{ii} \hat{y}_i + y_i h_{ii}^2 \right)$$

$$= \frac{1}{1-h_{ii}} (y_i - \hat{y}_i) = \frac{\hat{e}_i}{1-h_{ii}}$$

Q. E. D.

Budeme potřebovat podobné formule pro $\hat{\beta} - \hat{\beta}_{(-i)}$ a $SSE_{(-i)}$

Věta: 1) Nechť $\hat{\beta}_{(-i)}$ značí LSE par. β v modelu bez i-tého pozorování.

Potom platí

$$\underline{\hat{\beta} - \hat{\beta}_{(-i)} = \frac{(X^TX)^{-1} x_i \hat{e}_i}{1-h_{ii}} = (X^TX)^{-1} x_i \hat{e}_{(-i)}}$$

2) Pro součet residuálních čtverců $SSE_{(-i)}$ v modelu bez i-tého poz. platí

$$\underline{SSE_{(-i)} = \sum_{j=1}^n \hat{e}_j^2 - \frac{\hat{e}_i^2}{1-h_{ii}}} .$$

Dk: 1) stejně jako v dk. předchozí věty platí

$$\underline{\hat{\beta}_{(-i)} = S_1 + \frac{1}{1-h_{ii}} S_2} , \quad \text{kde} \quad S_1 = \hat{\beta} - y_i (X^TX)^{-1} x_i$$

a $S_2 = (X^TX)^{-1} x_i \hat{y}_i - y_i (X^TX)^{-1} x_i h_{ii}$

tedy

$$\hat{\beta} - \hat{\beta}_{(-i)} = y_i (X^T X)^{-1} x_i - \frac{1}{1-h_{ii}} \left((X^T X)^{-1} x_i \hat{y}_i - y_i (X^T X)^{-1} x_i h_{ii} \right) =$$

$$= (X^T X)^{-1} x_i \left(y_i - \frac{\hat{y}_i - y_i h_{ii}}{1-h_{ii}} \right) = (X^T X)^{-1} x_i \left(\frac{y_i - y_i h_{ii} - \hat{y}_i + y_i h_{ii}}{1-h_{ii}} \right)$$

$$= (X^T X)^{-1} x_i \underbrace{\left(\frac{y_i - \hat{y}_i}{1-h_{ii}} \right)}_{= \frac{\hat{e}_i}{1-h_{ii}} = \hat{e}_{(i)}}$$

2) $\underline{SSE_{(-i)}} = (y_{(-i)} - X_{(-i)}^T \hat{\beta}_{(-i)})^T (y_{(-i)} - X_{(-i)}^T \hat{\beta}_{(-i)}) =$

$$= \sum_{\substack{j=1 \\ j \neq i}}^{n} (y_j - x_j^T \hat{\beta}_{(-i)})^2 = \underline{\sum_{j=1}^{n} (y_j - x_j^T \hat{\beta}_{(-i)})^2 - (y_i - x_i^T \hat{\beta}_{(-i)})^2}$$

z bodu 1) $\hat{\beta}_{(-i)} = \hat{\beta} - \frac{(X^T X)^{-1} x_i \hat{e}_i}{1-h_{ii}}$ tzn.

$$SSE_{(-i)} = \sum_{j=1}^{n} \left(y_j - x_j^T \hat{\beta} + \frac{x_j^T (X^T X)^{-1} x_i \hat{e}_i}{1-h_{ii}} \right)^2 - \left(y_i - x_i^T \hat{\beta} + \frac{x_i^T (X^T X)^{-1} x_i \hat{e}_i}{1-h_{ii}} \right)^2$$

protože $x_j^T (X^T X)^{-1} x_i = h_{ij}$, dostaneme

$$SSE_{(-i)} = \sum_{j=1}^{n} \left(\hat{e}_j + \frac{h_{ij} \hat{e}_i}{1-h_{ii}} \right)^2 - \left(\hat{e}_i + \frac{h_{ii} \hat{e}_i}{1-h_{ii}} \right)^2 = \underbrace{\sum_{j=1}^{n} \left(\hat{e}_j + \frac{h_{ij} \hat{e}_i}{1-h_{ii}} \right)^2}_{A} - \frac{\hat{e}_i^2}{(1-h_{ii})^2}$$

$$A = \sum_{j=1}^{n} \hat{e}_j^2 + \frac{2\hat{e}_i}{1-h_{ii}} \underbrace{\sum_{j=1}^{n} h_{ij} \hat{e}_j}_{=0} + \frac{\hat{e}_i^2}{(1-h_{ii})^2} \underbrace{\sum_{j=1}^{n} h_{ij}^2}_{= h_{ii}}$$

protože $\hat{y} = Hy \Rightarrow H\hat{y} = H^2 y = Hy$ a tedy

$H\hat{e} = H(y - \hat{y}) = Hy - H\hat{y} = \underline{0}$

tedy $SSE_{(-i)} = \sum_{j=1}^{n} \hat{e}_j^2 + \frac{\hat{e}_i^2}{(1-h_{ii})^2} (h_{ii} - 1) = \sum_{j=1}^{n} \hat{e}_j^2 - \frac{\hat{e}_i^2}{1-h_{ii}}$.

Q.E.D.

Důsledek: V modelu (**) s $m+1$ parametry β a bez i-tého pozorování platí

$$E(SSE_{(-i)}) = (n-m-2)\,\sigma^2$$

takže
$$\hat{\sigma}^2_{(-i)} = \frac{SSE_{(-i)}}{n-m-2}$$
je nestranný odhad σ^2.

dále
$$\hat{\sigma}^2_{(-i)} = \frac{(1-h_{ii})(n-m-1)s_n^2 - \hat{e}_i^2}{(1-h_{ii})(n-m-2)} = \frac{1}{n-m-2}\left(SSE - \frac{\hat{e}_i^2}{1-h_{ii}}\right)$$

kde $s_n^2 = \frac{1}{n-m-1} SSE$ (pro plný model).

Dk: protože $E\hat{e}_i^2 = \operatorname{Var} \hat{e}_i = \sigma^2(1-h_{ii})$ dostaneme dle předchozí věty

$$E(SSE_{(-i)}) = \sum_{j=1}^{n} \sigma^2(1-h_{jj}) - \sigma^2 = \sigma^2\Big[(n-1) - \underbrace{\sum_{j=1}^{n} h_{jj}}_{=\operatorname{tr} H = m+1}\Big] = \sigma^2(n-m-2)$$

$$\hat{\sigma}^2_{(-i)} = \frac{1}{n-m-2} SSE_{(-i)} = \frac{1}{n-m-2}\Big(\underbrace{\sum_{j=1}^{n}\hat{e}_j^2}_{=SSE = (n-m-1)s_n^2} - \frac{\hat{e}_i^2}{1-h_{ii}}\Big) = \frac{1}{n-m-2}\,\frac{(1-h_{ii})SSE - \hat{e}_i^2}{1-h_{ii}}.$$

Q.E.D.

Pozn: dá se ukázat, že $SSE_{(-i)}$ a $\hat{e}_i$ jsou nezávislé n. vel.

protože $\frac{SSE_{(-i)}}{\sigma^2} \sim \chi^2(n-m-2)$ a $\frac{\hat{e}_i}{\sigma\sqrt{1-h_{ii}}} \sim N(0,1)$

dostaneme
$$\underbrace{\frac{\hat{e}_i}{\hat{\sigma}_{(-i)}\sqrt{1-h_{ii}}}}_{=\hat{t}_i} \sim t(n-m-2)$$

Tvrzení: Uvažujme model (**), kde $h(X) = m+1$ a $e \sim N_n(0, \sigma^2 I_n)$. Nechť pro $i \in \hat{n}$ platí $h_{ii} \neq 1$. Potom i-té ext. stud. reziduum $\hat{t}_i \sim t(n-m-2)$

POZN: $\hat{t}_i$ lze použít pro test hypotézy, zda je i-té pozorování odlehlé (outlier)

H_0: i-té pozorování je není odlehlé v modelu M

$\times$

H_1: i-té pozorování je odlehlé v M

kde odlehlé znamená odlehlé vzhledem k M: $Y_i \sim N_n(X\beta, \sigma^2 I_n)$:

a) střední hodnota i-tého poz. se nerovná té dané modelem

b) pozorovaná hodnota Y_i je neobvyklá za platnosti M

H_0 zamítneme pokud $|\hat{t}_i| > t_{1-\frac{\alpha}{2}}(n-m-2)$

$\approx u_{1-\frac{\alpha}{2}} \doteq 2$ pro $\alpha = 0{,}05$ a n velké

- pokud test použijeme na všechna pozorování, je třeba aplikovat nějakou korekci na vícenásobné testování, např. Bonferroni

$\hat{t}_i \in (-2, +2)$ OK

POZN: vztah $\hat{e}_{(-i)}$ a $\hat{t}_i$

$$\hat{e}_{(-i)} = \frac{\hat{e}_i}{1-h_{ii}} \Rightarrow \quad E\,\hat{e}_{(-i)} = 0 \quad \operatorname{Var} \hat{e}_{(-i)} = \frac{\sigma^2}{1-h_{ii}}$$

standardizované PRESS residuum $\dfrac{\hat{e}_{(-i)}}{\sqrt{\operatorname{Var}(\hat{e}_{(-i)})}} = \dfrac{\frac{\hat{e}_i}{1-h_{ii}}}{\frac{\sigma}{\sqrt{1-h_{ii}}}} = \dfrac{\hat{e}_i}{\sigma\sqrt{1-h_{ii}}} = r_i$

pokud použijeme $\hat{\sigma}^2_{(-i)}$ jako odhad σ^2, pak <u>studentizované PRESS residuum</u> $\dfrac{\hat{e}_i}{\hat{\sigma}_{(-i)}\sqrt{1-h_{ii}}} = \underline{\hat{t}_i}$

POZN: • $\hat{e}_{(-i)} = \dfrac{\hat{e}_i}{1-h_{ii}} \Rightarrow$ pokud i-té pozorování má velký potenciál h_{ii}, bude $\hat{e}_{(-i)}$ mnohem větší než $\hat{e}_i$, pozorování s velkým h_{ii} jsou dobře modelována, ale měřeno $\hat{e}_{(-i)}$ mohou špatně predikovat 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.