IS a t-testy pro parametry

- pokud se model ukáže jako významný, bude nás zajímat, které koeficienty přispívají
- lze použít IS a TH stejně jako v jednorozměrné regresi
- výsledky jsou odvozeny pro normální chyby
- v praxi se používají i pro jiné typy chyb
  (za jistých předpokladů budou platit asymptoticky, lze si pomoci pro velká n)

pro konstrukci použijeme dokázanou vlastnost

$$T_j = \frac{\hat{\beta}_j - \beta_j}{s_n \sqrt{v_j}} \sim t(n-m-1), \quad \text{kde} \quad v_j = (X^T X)^{-1}_{jj}$$

standardním postupem získáme 100(1−α)% IS pro β_j ve tvaru

$$\left(\hat{\beta}_j - t_{1-\frac{\alpha}{2}}(n-m-1)\, s_n \sqrt{v_j},\ \hat{\beta}_j + t_{1-\frac{\alpha}{2}}(n-m-1)\, s_n \sqrt{v_j} \right)$$

s jejich pomocí lze odvodit kritický obor pro test

$$H_0: \beta_j = b_j \quad \times \quad H_1: \beta_j \neq b_j$$

(H_0 zamítneme pokud $b_j \notin (\underline{\beta}_j, \overline{\beta}_j)$, kde IS pro β_j)

ve tvaru

$$\frac{|\hat{\beta}_j - b_j|}{s_n \sqrt{v_j}} > t_{1-\frac{\alpha}{2}}(n-m-1)$$

pro $b_j = 0$ dostaneme test významnosti β_j, tzn. $H_0: \beta_j = 0$ zamítneme, pokud

$$\frac{|\hat{\beta}_j|}{s_n \sqrt{v_j}} > t_{1-\frac{\alpha}{2}}(n-m-1)$$

POZN.
- pokud nejsou porušeny předpoklady modelu nebo není přítomna kolinearita, lze zvážit odstranění všech nevýznamných proměnných (dle t-testu)
- v případě kolinearity, model může být významný (dle celkového F-testu), ale všechny nebo téměř všechny proměnné se mohou jevit jako nevýznamné (dle t-testu)
- naopak, pokud má model velký počet možných proměnných, některé proměnné se mohou jevit významné, i když jsou náhodným šumem
- při použití t-testů je třeba být obezřetný (postupovat obezřetně)

Př. 5.26, str. 230

Př. 5.27, str. 231

POZN: statistiky F, R^2 a t jsou užitečné pro rozhodnutí efektivity jedn. proměnných, nemohou být ale používány výhradně automaticky

Obecná lineární hypotéza

F-test (celkový) a t-testy jsou spec. případem obecné lineární hypotézy

$$H_0 : \mathbb{C}\beta = b \qquad \times \qquad H_1 : \mathbb{C}\beta \neq b,$$

kde $\mathbb{C} \in \mathbb{R}^{r \times (m+1)}$ a $h(\mathbb{C}) = r$ (tzn. $r \leq m+1$)

- rovnice $\mathbb{C}\beta = b$ reprezentují r lin. nezávislých podmínek

$$\sum_{j=0}^{m} c_{ij} \beta_j = b_i \qquad i = 1, \dots, r$$

POZN: a) volba $b = (0, \dots, 0)^T$ a $\mathbb{C} = \begin{pmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & \cdots \\ \vdots & & & & \\ 0 & 0 & \cdots & 0 & 1 \end{pmatrix}_{m \times (m+1)}$

vede na test

$$H_0 : \mathbb{C}\beta = 0 \iff H_0 : \beta_1 = \beta_2 = \cdots = \beta_m = 0$$

b) volba $b = 0$ a $\mathbb{C} = (0, \dots, 0, \underbrace{1}_{j}, 0, \dots 0)$ vede na test

$$H_0 : \beta_j = 0$$

c) v modelu $Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + e$

chceme testovat zároveň, že $\beta_2 = 0$ a $\beta_3 = \beta_4$

to lze udělat volbou $\mathbb{C} = \begin{pmatrix} 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & -1 \end{pmatrix}$ $b = (0, 0)^T$

Pro test H_0 uvažujeme 2 modely:

1) plný model (full model) – bez podmínek na $\mathbb{C}\beta$
2) redukovaný model (reduced model) – za předpokladu, že platí $H_0 : \mathbb{C}\beta = b$

Označme příslušné residuální součty čtverců $\underline{SSE_F}$ a $\underline{SSE_R}$

(bude platit $SSE_F \le SSE_R$)

- pokud neplatí H_0, dá se očekávat, že $\Delta SSE = SSE_R - SSE_F$ bude významně větší než rozptyl chyb σ^2, H_0 tedy budeme zamítat, pokud $\dfrac{\Delta SSE}{s_n^2}$ bude velké

- obecný F-test
$$F = \frac{\frac{\Delta SSE}{r}}{s_n^2} \sim F(r, n-m-1)$$
ukázáno pro normální chyby za platnosti H_0

__Př.__ uvažujme F-test pro $H_0: \beta_1 = \beta_2 = \dots = \beta_m = 0$ v modelu (**) (full model)

redukovaný model bude: $Y_i = \beta_0 + e_i$, $i = 1, \dots, n$ $\Rightarrow$ $\hat{\beta}_0 = \bar{y}$

a $SSE_R = \sum_{i=1}^{n} (y_i - \bar{y})^2 = SST$, tedy

$$\underline{\Delta SSE} = SST - SSE_F = \sum_{i=1}^{n} (y_i - \bar{y})^2 - \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = \underline{SSR}$$

a statistika
$$F = \frac{\frac{SSR}{m}}{s_n^2} = F_{overall} \sim F(m, n-m-1)$$
jak jsme již ukázali

__Věta:__ Nechť v modelu (**) platí, že $e_1, \dots, e_n$ jsou nezávislé a $e_i \sim N(0, \sigma^2)$ a $h(X) = m+1$. Označme SSE_F residuální s.č. plného modelu a SSE_R rez. s.č. ~~modelu~~ modelu, kde platí $\underline{H_0: C\beta = b}$. Potom, za platnosti H_0 je splněno:

$$\underline{F = \frac{\frac{\Delta SSE}{r}}{s_n^2} \sim F(r, n-m-1)}.$$

Lemma: Označme $\hat{\beta}_F$ a $\hat{\beta}_R$ LSE par. β v plném a redukovaném modelu. Potom platí

1) $\hat{\beta}_F = \hat{\beta}_R - (X^TX)^{-1}C^TA(C\hat{\beta}_F - b)$, kde $A = (C(X^TX)^{-1}C^T)^{-1}$.

2) $\Delta SSE = SSE_R - SSE_F = (C\hat{\beta}_F - b)^T A (C\hat{\beta}_F - b)$.

Dk. 1) víme $\hat{\beta}_F = (X^TX)^{-1}X^Ty$, musíme najít $\hat{\beta}_R$

budeme minimalizovat $g(\beta) = (y - X\beta)^T(y - X\beta)$ za podmínky $C\beta = b$

Lagrangeova funkce

$$L = L(\beta, \lambda) = g(\beta) - 2\lambda^T(C\beta - b), \quad \text{kde} \quad \lambda = (\lambda_1, \dots, \lambda_r)$$

$$L = y^Ty - 2\underbrace{y^TX\beta}_{= \beta^TX^Ty} + \beta^TX^TX\beta - 2\underbrace{\lambda^TC\beta}_{= \beta^TC^T\lambda} + 2\lambda^Tb$$

a tedy

$$\frac{\partial L}{\partial \beta} = \left(\frac{\partial L}{\partial \beta_0}, \frac{\partial L}{\partial \beta_1}, \dots, \frac{\partial L}{\partial \beta_m}\right)^T = 2X^TX\beta - 2X^Ty - 2C^T\lambda = 0$$

$$\frac{\partial L}{\partial \lambda} = \left(\frac{\partial L}{\partial \lambda_1}, \dots, \frac{\partial L}{\partial \lambda_r}\right)^T = C\beta - b = 0$$

z první rovnice $\hat{\beta}_R = (X^TX)^{-1}X^Ty + (X^TX)^{-1}C^T\lambda = \hat{\beta}_F + (X^TX)^{-1}C^T\lambda$

dosadíme do druhé:

$$C\hat{\beta}_R - b = C\hat{\beta}_F - b + C(X^TX)^{-1}C^T\lambda = 0$$

tzn. $\lambda = -(C(X^TX)^{-1}C^T)^{-1}(C\hat{\beta}_F - b)$

$$\hat{\beta}_R = \hat{\beta}_F - (X^TX)^{-1}C^T\underbrace{(C(X^TX)^{-1}C^T)^{-1}}_{= A}(C\hat{\beta}_F - b) = \hat{\beta}_F - (X^TX)^{-1}C^TA(C\hat{\beta}_F - b)$$

2) z dk. věty $\hat{\beta}=(X^TX)^{-1}X^TY$ víme, že

$$g(\beta) - g(\hat{\beta}_F) = (\beta-\hat{\beta}_F)^T X^TX(\beta-\hat{\beta}_F) \quad \forall \beta$$

dosadíme $\beta=\hat{\beta}_R$

$$\Delta SSE = g(\hat{\beta}_R)-g(\hat{\beta}_F) = (\hat{\beta}_R-\hat{\beta}_F)^T X^TX(\hat{\beta}_R-\hat{\beta}_F) =$$

$$= (C\hat{\beta}_F - b)^T A^T C \cancel{(X^TX)^{-1}X^TX} (XX)^{-1}C^T A (C\hat{\beta}_F - b) = *$$

protože $A^T=A$, platí $A^T \underbrace{C(XX)^{-1}C^T}_{A^{-1}} A = A$

$$* = (C\hat{\beta}_F - b)^T A (C\hat{\beta}_F-b)$$

Q.E.D

Dk věty: nejdříve ukážeme, že $\frac{\Delta SSE}{\sigma^2}\sim\chi^2(r)$ za $H_0: C\beta=b$

za H_0: $Y\sim N(X\beta_R, \sigma^2 I_n)$ a $\hat{\beta}_F=(X^TX)^{-1}X^TY$, tzn.

$$\hat{r} = C\hat{\beta}_F - b \sim N(E(\hat{r}), Cov(\hat{r}))$$

$$E\hat{r} = E(C\hat{\beta}_F - b) = E(C(X^TX)^{-1}X^TY) - b = C(X^TX)^{-1}X^T \underbrace{EY}_{=X\beta_R} - b =$$

$$= C(X^TX)^{-1}X^TX\beta_R - b = C\beta_R - b = 0 \text{ za } H_0$$

$$Cov(\hat{r}) = C\,Cov(\hat{\beta}_F)C^T = \sigma^2 \underbrace{C(X^TX)^{-1}C^T}_{=A^{-1}} = \sigma^2(A^{-1})$$

$$\Rightarrow \hat{r} = C\hat{\beta}_F - b \sim N(0,\sigma^2A^{-1}) \text{ za } H_0$$

tzn. $\frac{\Delta SSE}{\sigma^2} = \frac{\hat{r}^T A \hat{r}}{\sigma^2}\sim\chi^2(r)$ (bod 4) věty str. 55) $Z\sim N_r(0,\Sigma) \Rightarrow Z^T\Sigma^{-1}Z\sim\chi^2(r)$)

bod 1) $\Rightarrow$ β_F a D_n^2 nezávislé

ΔSSE je funkcí pouze $\hat{\beta}_F$, tzn. nezávisí na s_n^2

takže

$$F = \frac{\dfrac{\Delta SSE}{\sigma^2 \cdot r}}{\dfrac{(n-m-1)s_n^2}{\sigma^2 (n-m-1)}} = \frac{\dfrac{\Delta SSE}{r}}{s_n^2} \sim F(r, n-m-1)$$

Q.E.D.

POZN: Praktický rozklad $SST = SSE + SSR$ dostaneme

$$\Delta SSE = SSR_F - SSR_R$$

interpretace: nárůst regresního součtu čtverců díky neplatnosti H_0

dále

$$SSR_F = SSR_R + \Delta SSE$$

extra sum of squares přidaná k SSR díky neplatnosti H_0

Např. pokud $\beta = (\beta_0, \beta_1, \dots, \beta_{m-1}, 0)$ tzn. $\beta_m = 0$

a skutečný model má $\beta = \beta_F$, potom ΔSSE je extra regresní součet čtverců získaný díky přidání β_m do modelu.

- umožňuje rozklad SSR plného modelu na jednotlivé části (x_1, $x_2|x_1$, $x_3|x_2 x_1$, ...)

Př. analogie k Př. 5.25 str. 238

POZN: joint confidence regions viz Ex. 5.30 str. 233 ???

Predikce

jakmile máme adekvátní model, můžeme ho použít pro bodové a intervalové predikce jako v jednorozměrné regresi

a) predikce EY_{x}

nechť $x_0 = (1, x_{0,1}, \dots, x_{0,m})^T$ je nový bod proměnné x

bodový odhad $E(Y_{x_0})$ je $\underline{\hat{y}_{x_0} = \hat{\beta}_0 + \sum_{j=1}^{m} x_{0,j}\hat{\beta}_j = x_0^T\hat{\beta}}$

tzn. $\mathrm{Var}(\hat{Y}_{x_0}) = x_0^T \,\mathrm{Var}\,\hat{\beta}\, x_0 = \underbrace{\sigma^2 x_0^T (X^TX)^{-1} x_0}_{*}$ a může být odhadnut

pomocí $\hat{\sigma}^2(\hat{Y}_{x_0}) = s_n^2 \left[x_0^T (X^TX)^{-1} x_0 \right]$

rozptyl predikce

spec. případ $\underline{x_0^T = x_i^T}$ (i-tý řádek matice X)

$\hat{\sigma}^2(\hat{Y}_{x_i}) = s_n^2 \left[x_i^T (X^TX)^{-1} x_i \right] = \underline{s_n^2 h_{ii}}$, kde $h_{ii} = (H)_{ii}$

$H = X(X^TX)^{-1}X^T$

pro normální chyby lze odvodit IS pro $EY_{x_0} = \mu_{x_0}$

protože $\hat{Y}_{x_0}$ je LK náh. vel. s vícerozměrným norm. rozdělením, má normální rozdělení se $E\hat{Y}_{x_0} = \mu_{x_0} = x_0^T\beta$ a $\mathrm{Var}\,\hat{Y}_{x_0} = *$

tzn. $\dfrac{\hat{Y}_{x_0} - \mu_{x_0}}{\sigma\sqrt{x_0^T(X^TX)^{-1}x_0}} \sim N(0,1)$ a díky nezávislosti $\hat{\beta}$ a s_n^2

$\dfrac{\hat{Y}_{x_0} - \mu_{x_0}}{s_n\sqrt{x_0^T(X^TX)^{-1}x_0}} \sim t(n-m-1)$ $\Rightarrow$ $\underline{100(1-\alpha)\%\ \text{IS pro } \mu_{x_0} \text{ je}}$

$$\left(\hat{Y}_{x_0} \pm t_{1-\frac{\alpha}{2}}(n-m-1) \cdot s_n \sqrt{x_0^T(X^TX)^{-1}x_0} \right)$$

b) intervaly predikce pro Y_{x_0}

bodový odhad je opět $\hat{Y}_{x_0}$

pokud Y_{x_0} je skutečná hodnota Y_x v bodě $x = x_0$, potom Y_{x_0} a $\hat{Y}_{x_0}$ budou nezávislé za předpokladu, že poz. $Y_{x_0}, Y_1, \dots, Y_n$ jsou nezávislé

(což předpokládáme)

potom

$$\mathrm{Var}(\hat{Y}_{x_0} - Y_{x_0}) = \mathrm{Var}(\hat{Y}_{x_0}) + \mathrm{Var}(Y_{x_0}) = \sigma^2(1 + x_0^T(X^TX)^{-1}x_0),$$

tudíž

$$\frac{\hat{Y}_{x_0} - Y_{x_0}}{\sigma\sqrt{1 + x_0^T(X^TX)^{-1}x_0}} \sim N(0,1) \quad \text{a} \quad \frac{\hat{Y}_{x_0} - Y_{x_0}}{s_n\sqrt{1 + x_0^T(X^TX)^{-1}x_0}} \sim t(n-m-1)$$

za předpokladu normality chyb

100(1-α)% IP pro Y_{x_0} tedy je

$$\left(\hat{Y}_{x_0} \pm t_{1-\frac{\alpha}{2}}(n-m-1) \cdot s_n\sqrt{1 + x_0^T(X^TX)^{-1}x_0}\right)$$

Př.

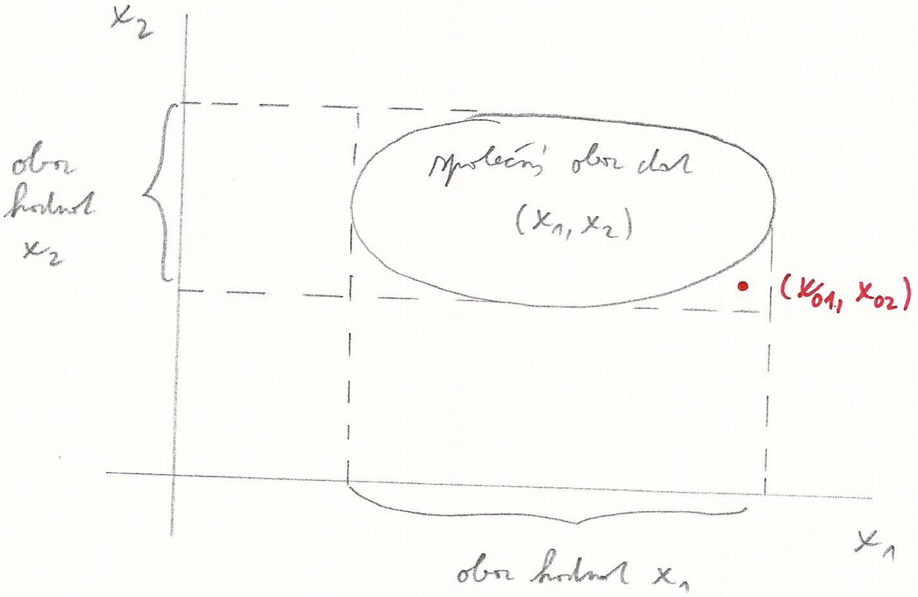
Pozn.: extrapolace

- v jednoduché LR kvalita predikce závisela na vzdálenosti x_0 od $\bar{x}$
- je třeba si dát pozor na predikce mimo $\langle x_{min}, x_{max} \rangle$
- podobné závěry platí i pro vícerozměrnou LR
- protože rozptyl predikce je úměrný $x_0^T(X^TX)^{-1}x_0$, v ~~bodech~~ bodech s velkými hodnotami této veličiny nebude predikce spolehlivá
- spec. pokud x_i^T jsou pozorovaná data, můžeme očekávat, že body s největšími hodnotami $x_i^T(X^TX)^{-1}x_i = h_{ii}$ budou na hranici množiny, kde je predikce spolehlivá

tzn., že vnitřek elipsoidu

$$x^T (X^T X)^{-1} x \le \max_{1 \le i \le n} h_{ii}$$

může být považován za přípustný obor pro predikce



(x_{01}, x_{02}) leží uvnitř oboru hodnot pro oba x_1 i x_2 ale vně společného oboru původních dat