Gauss - Markov theorem

e_i i.i.d. $N(0,\sigma^2)$ $\Rightarrow$ OLS $\hat\beta$ je MLE, tzn. je eficientní MVUE pro β

chyby nenormální - ukážeme, že OLS $\hat\beta$ je BLUE (best linear unbiased est.) parametru β (za jistých podmínek)

- mohou ale ex. lepší lineární vychýlené odhady nebo nelineární odhady

Def. Nechť β je vektor regresních parametrů v LM. Řekneme, že $\hat\beta$ je lineární odhad β, jestliže každé $\hat\beta_i$ je LK pozorování Y_i, $i=1,\dots,n$, tedy

$$\hat\beta_i = \sum_{j=1}^{n} a_{ij} Y_j \qquad i = 0,\dots,m$$

v maticovém zápisu

$$\hat\beta = AY, \text{ kde } A = (a_{ij}) \qquad \begin{matrix} i=0,\dots,m \\ j=1,\dots,n \end{matrix}$$

Pozn: Protože v modelu $Y = X\beta + e$ platí $h(X) = m+1$, potom OLS $\hat\beta$ je lineární, neboť $\hat\beta = \underbrace{(X^TX)^{-1}X^T}_{A} Y$.

Věta (Gauss - Markov)
Uvažujme model $Y = X\beta + e$, kde matice X má plnou hodnost, e_i, $i=1,\dots,n$, jsou nekorelované a $e_i \sim (0,\sigma^2)$. Potom OLS odhad $\hat\beta$ je BLUE parametru β.

Dk: nechť $\hat\beta = AY$ je lineární odhad β, aby byl nestranný musí platit

$E\hat\beta = \beta$, tzn. $E(AY) = A\,EY = AX\beta = \beta$, tedy $(AX - I_{m+1})\beta = 0$

protože to musí platit $\forall \beta \in \mathbb{R}^{m+1}$, dostáváme $AX - I_{m+1} = 0$, nebo ekv. $AX = I_{m+1}$

spočteme kovarianční matici $\hat{\beta}$

$$\mathrm{Cov}(\hat{\beta}) = \mathrm{Cov}(AY) = A \underbrace{\mathrm{Cov}\, Y}_{=\sigma^2 I_n} A^T = \sigma^2 A A^T$$

zapišme A ve tvaru $A = (X^TX)^{-1}X^T + D$

(D je rozdíl mezi A a maticí pro OLS odhad)

pokud ukážeme, že pro nestranný lin. odhad $\hat{\beta} = AY$, který minimalizuje rozptyl, musí platit $D = 0$, bude věta dokázána

dosazením dostaneme

$$\mathrm{Cov}(\hat{\beta}) = \sigma^2((X^TX)^{-1}X^T + D)(X(X^TX)^{-1} + D^T) =$$

$$= \sigma^2 \left[(X^TX)^{-1} + DX(X^TX)^{-1} + (X^TX)^{-1}X^TD^T + DD^T \right]$$

z podmínky nestrannosti

$$AX = [(X^TX)^{-1}X^T + D]X = I_{m+1} + DX = I_{m+1} \Rightarrow DX = 0 \text{ a}$$

tedy i $X^TD^T = 0$

tzn. $\mathrm{Cov}(\hat{\beta}) = \sigma^2 [(X^TX)^{-1} + DD^T]$

pro diagonální prvky platí

$$\mathrm{Var}\, \hat{\beta}_i = \sigma^2 \left[v_i + \sum_{j=1}^{n} d_{ij}^2 \right] \qquad i = 0, \dots, m$$

protože $v_j \geq 0$ a $\sum_{j=1}^{n} d_{ij}^2 \geq 0$ $\Rightarrow$ $\mathrm{Var}\, \hat{\beta}_i$ je minimalizován volbou

$\sum_{j=1}^{n} d_{ij}^2 = 0$, tj. $d_{ij} = 0$ $\quad j = 1, \dots, n$

platí $\forall i = 0, \dots, m$ $\Rightarrow$ $D = 0$

tzn. lin. nestranný odhad $\hat{\beta}$, který minimalizuje $\mathrm{Var}(\hat{\beta}_i)$, $i = 0, \dots, m$,

je $\hat{\beta} = (X^TX)^{-1}X^TY$.

Q.E.D.

Testování modelu – tabulka ANOVA

a) Celkový F-test (overall F-test)

- zajímá nás, zda je model stat. významný, tj. zda alespoň jeden z koef. $\beta_1, \dots, \beta_m$ je nenulový
- mohli bychom testovat jednotlivé koef. $H_0: \beta_j = 0$ pomocí alternativy t-testu
- celková chyba I. druhu by takto ale mohla* být velká, pokud máme hodně proměnných

  *museli bychom hodně snížit α pro jednotlivé testy, což zvýší pravděpodobnost chyby II. druhu (tzn. rizika akceptování nenulových koef. jako nulových a tedy vynechání významných proměnných z modelu)
- navíc je zde problém multikolinearity (viz později)

  jejíž jedním efektem jsou velké stand. chyby odhadu, to může vést k akceptování všech koeficientů jako 0, i když je model celkově významný (uvidíme na cv.)

Bylo by dobré mít jednu statistiku pro test

$$\underline{H_0: \beta_1 = \beta_2 = \dots = \beta_m = 0} \quad \times \quad H_1: (\exists i \in \hat{m},\ \beta_i \neq 0)$$

alespoň jeden $\beta_i \neq 0$, $i = 1, \dots, m$

ANOVA přístup pro jedn. regresi naznačuje, že statistika

$$F = \frac{\frac{SSR}{m}}{s_r^2}$$

by mohla být užitečná

(vyplyne i z obecnějších přístupů k testování později)

Označení: $\bar{x}_j = \frac{1}{n}\sum_{i=1}^{n} x_{ij}$ průměr j-tého sloupce matice $\mathbb{X}$

$$\bar{\mathbb{X}} = \begin{pmatrix} \bar{x}_0 & \bar{x}_1 & \cdots & \bar{x}_m \\ \bar{x}_0 & \bar{x}_1 & & \vdots \\ \vdots & \vdots & & \vdots \\ \bar{x}_0 & \bar{x}_1 & & \bar{x}_m \end{pmatrix}_{n \times m+1}$$

$$(\mathbb{X}_c)_{ij} = x_{ij} - \bar{x}_j \qquad i = 1, \dots, n, \quad j = 1, \dots, m$$

$\uparrow$ centrované matice regrese

Věta: V modelu $\mathbb{Y} = \mathbb{X}\beta + \varepsilon$ (**), kde e_i jsou nekorelované a $e_i \sim (0, \sigma^2)$, $i = 1, \dots, n$, platí

$$E\left(\frac{SSR}{n}\right) = \sigma^2 + \frac{\beta^T(\mathbb{X}-\bar{\mathbb{X}})^T(\mathbb{X}-\bar{\mathbb{X}})\beta}{n} = \sigma^2 + \frac{\beta_1^T \mathbb{X}_c^T \mathbb{X}_c \beta_1}{n},$$

kde $\beta_1 = (\beta_1, \dots, \beta_m)$.

Dk:
$$\hat{y}_i = \hat{\beta}_0 + \sum_{j=1}^{m} x_{ij}\hat{\beta}_j, \qquad \frac{\partial SSE}{\partial \beta_0} = \sum_{i=1}^{n}\left(y_i - \left(\beta_0 + \sum_{j=1}^{m} x_{ij}\hat{\beta}_j\right)\right) = 0$$

$$\Downarrow$$

$$\hat{\beta}_0 = \bar{y} - \sum_{j=1}^{m} \bar{x}_j \hat{\beta}_j$$

odtud
$$\underline{\hat{y}_i - \bar{y} = \sum_{j=1}^{m}(x_{ij} - \bar{x}_j)\hat{\beta}_j} \qquad i = 1, \dots, n$$

zapsáno maticově $\underline{\hat{\mathbb{Y}} - \bar{\mathbb{Y}} = (\mathbb{X} - \bar{\mathbb{X}})\hat{\beta}}$, kde $\bar{\mathbb{Y}} = (\bar{y}, \bar{y}, \dots, \bar{y})^T_{1\times n}$

protože první sloupec matice $\mathbb{X} - \bar{\mathbb{X}}$ je nulový

$$\underline{SSR} = \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2 = (\hat{\mathbb{Y}} - \bar{\mathbb{Y}})^T(\hat{\mathbb{Y}} - \bar{\mathbb{Y}}) = \hat{\beta}^T \underbrace{(\mathbb{X}-\bar{\mathbb{X}})^T(\mathbb{X}-\bar{\mathbb{X}})}_{A}\hat{\beta} = \underline{\hat{\beta}^T A \hat{\beta}}$$

V: $Z = \mathbb{Y}^T A \mathbb{Y}$ je kvadratická forma a nechť $E\mathbb{Y} = \mu$ a $\operatorname{Cov}\mathbb{Y} = \Sigma$.

Potom platí $\underline{EZ = \operatorname{tr}(A\Sigma) + \mu^T A \mu}$

nejdříve zjednodušíme matici A

$$\bar{X} = \frac{1}{n} B X, \quad \text{kde} \quad B = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ \vdots & & & \\ 1 & 1 & \cdots & 1 \end{pmatrix}_{n \times n} \quad \text{a tedy}$$

$$X - \bar{X} = \left(I_n - \frac{1}{n} B\right) X \qquad (X - \bar{X})^T = X^T \left(I_n - \frac{1}{n} B\right)$$

$$A = (X-\bar{X})^T (X - \bar{X}) = X^T \underbrace{\left(I_n - \frac{1}{n} B\right)^2}_{\substack{\| \\ I_n - \frac{2}{n} B + \frac{B^2}{n^2}}} X = X^T \left(I_n - \frac{1}{n} B\right) X$$

$$\frac{B^2}{n^2} = \frac{B}{n}$$

dále $$\underset{\substack{\| \\ \operatorname{cov} \hat{\beta}}}{A \Sigma} = \sigma^2 X^T \left(I_n - \frac{1}{n} B\right) X (X^T X)^{-1}$$

$$\operatorname{tr}(A\Sigma) = \sigma^2 \operatorname{tr}\Big\{ \underbrace{X (X^T X)^{-1} X^T}_{H} \left(I_n - \frac{1}{n} B\right) \Big\} = \sigma^2 \operatorname{tr}\Big\{ H - \frac{1}{n} H B \Big\}$$

$\operatorname{tr}(AB) = \operatorname{tr}(BA)$

$$= \sigma^2 \left(\operatorname{tr} H - \frac{1}{n} \operatorname{tr}(HB) \right) = \sigma^2 \Big(\underset{m+1}{\operatorname{tr} H} - \frac{1}{n} \underset{n}{\operatorname{tr} B} \Big) = \sigma^2 m$$

víme $HX = X$ a $\mathbb{1} = (1, \dots, 1)^T$ je první sloupec X, tedy $H \mathbb{1} = \mathbb{1}$

a tedy $HB = B$

celkem tedy

$$E\left(\frac{SSR}{m}\right) = \frac{1}{m} \left(\sigma^2 m + \beta^T (X - \bar{X})^T (X - \bar{X}) \beta \right) = \sigma^2 + \frac{1}{m} \beta^T (X - \bar{X})^T (X - \bar{X}) \beta$$

dále $(X - \bar{X}) \beta = X_c \beta_1$ protože první sloupec matice $X - \bar{X}$ je nulový vektor.

Q.E.D.

POZN: pokud $\beta_1 = 0$ potom $E \frac{SSR}{m} = \sigma^2 = E\, s_n^2$

Pokud $\beta_1 \neq 0$ vyplývá, že $E \frac{SSR}{m} > \sigma^2$, tedy velké hodnoty

$$F = \frac{\frac{SSR}{m}}{s_n^2}$$ budou znamenat zamítnutí $H_0: \beta_1 = 0$

budeme potřebovat rozdělení F za H_0.

Věta: Nechť v modelu $\mathbb{Y} = \mathbb{X}\beta + \boldsymbol{e}$ (**) jsou $e_1, \dots, e_n$ i.i.d. $N(0, \sigma^2)$. Pokud $\beta_D = 0$ $(\beta_1 = \beta_2 = \dots = \beta_m = 0)$, potom

$$\underline{F \sim F(m, n-m-1)}$$

Dk. v důkazu minulé věty jsme ukázali

$$SSR = \hat{\beta}^T \mathbb{X}^T \left(I_n - \tfrac{1}{n}\mathbb{B}\right)\mathbb{X}\hat{\beta} = \hat{\mathbb{Y}}^T \left(I_n - \tfrac{1}{n}\mathbb{B}\right)\hat{\mathbb{Y}}$$

potřebujeme zjednodušit

$$\hat{\mathbb{Y}} = \mathbb{H}\mathbb{Y} = \mathbb{H}(\mathbb{X}\beta + \boldsymbol{e}) = \mathbb{H}(\mathbb{1}\beta_0 + \mathbb{X}_D\beta_D + \boldsymbol{e}) = \beta_0 \underbrace{\mathbb{H}\mathbb{1}}_{=\mathbb{1}} + \underbrace{\mathbb{H}\mathbb{X}_D\beta_D}_{=0} + \mathbb{H}\boldsymbol{e}$$

$$= \beta_0 \mathbb{1} + \mathbb{H}\boldsymbol{e}$$

$$\underline{SSR} = (\beta_0 \mathbb{1}^T + \boldsymbol{e}^T\mathbb{H})\left(I_n - \tfrac{1}{n}\mathbb{B}\right)(\beta_0\mathbb{1} + \mathbb{H}\boldsymbol{e}) = \underline{\boldsymbol{e}^T\mathbb{H}\left(I_n - \tfrac{1}{n}\mathbb{B}\right)\mathbb{H}\boldsymbol{e}}$$

ale $\left(I_n - \frac{1}{n}\mathbb{B}\right)\mathbb{1} = 0$ a $\left(I_n - \frac{1}{n}\mathbb{B}\right)$ je symetrická

dále platí $\mathbb{H} = \mathbb{H}^T$, $\mathbb{H}^2 = \mathbb{H}$ a $\mathbb{H}\mathbb{B} = \mathbb{B}\mathbb{H} = \mathbb{B}$ (protože $\mathbb{H}\mathbb{1} = \mathbb{1}$)

celkem tedy dostaneme

$$\underline{SSR} = \boldsymbol{e}^T \underbrace{\left(\mathbb{H} - \tfrac{1}{n}\mathbb{B}\right)}_{\mathbb{C}} \boldsymbol{e} = \underline{\boldsymbol{e}^T \mathbb{C}\, \boldsymbol{e}}$$

pro matici $\mathbb{C}$ platí $\underline{\mathbb{C}^T} = \left(\mathbb{H}^T - \frac{1}{n}\mathbb{B}^T\right) = \left(\mathbb{H} - \frac{1}{n}\mathbb{B}\right) = \underline{\mathbb{C}}$

a $\underline{\mathbb{C}^2} = \left(\mathbb{H} - \frac{1}{n}\mathbb{B}\right)\left(\mathbb{H} - \frac{1}{n}\mathbb{B}\right) = \mathbb{H}^2 - \frac{1}{n}\mathbb{H}\mathbb{B} - \frac{1}{n}\mathbb{B}\mathbb{H} + \frac{1}{n^2}\mathbb{B}^2 =$

$$= \mathbb{H} - \tfrac{2}{n}\mathbb{B} + \tfrac{1}{n}\mathbb{B} = \mathbb{H} - \tfrac{1}{n}\mathbb{B} = \underline{\mathbb{C}}$$

tzn. $\mathbb{C}$ je symetrická a idempotentní, tedy

$$\underline{h(\mathbb{C})} = tr(\mathbb{C}) = tr\left(\mathbb{H} - \tfrac{1}{n}\mathbb{B}\right) = m + 1 - 1 = \underline{m}$$

věta o spektrálním rozkladu $\Rightarrow$ $\exists\, Q$ OG a diag Λ tak, že

$$C = Q^T \Lambda Q = Q^T \underbrace{\begin{pmatrix} I_m & 0 \\ 0 & 0 \end{pmatrix}}_{\substack{\text{idempotentní matice má} \\ \text{vlastní čísla 0 nebo 1}}} Q$$

$$\underline{SSR} = e^T Q^T \begin{pmatrix} I_m & 0 \\ 0 & 0 \end{pmatrix} \underbrace{Q e}_{Z \sim N(0, \sigma^2 I_n)} = Z^T \begin{pmatrix} I_m & 0 \\ 0 & 0 \end{pmatrix} Z = \underline{\sum_{i=1}^{m} Z_i^2} \quad \text{kde}$$

$\underline{Z_i \sim N(0, \sigma^2) \text{ nezávislé}}$

$$\frac{Z_i}{\sigma} \sim N(0,1) \quad \text{a} \quad \frac{SSR}{\sigma^2} \sim \chi^2(m)$$

$$\text{tzn.} \quad \frac{\overbrace{\frac{SSR}{\sigma^2 \cdot m}}^{\chi^2(m)/m}}{\underbrace{\frac{(n-m-1)\, s_n^2}{\sigma^2}}_{\chi^2(n-m-1)} \cdot \frac{1}{n-m-1}} = \frac{\frac{SSR}{m}}{s_n^2} = F \sim F(m, n-m-1)$$

pokud ukážeme, že SSR a s_n^2 jsou nezávislé

stačí ukázat, že SSR je nezávislé na reziduích $\hat{e}_i$, $i = 1, \dots, n$

$$SSR = e^T H \left(I_n - \tfrac{1}{n} B\right) H e = e^T H \underbrace{\left(I_n - \tfrac{1}{n} B\right)\left(I_n - \tfrac{1}{n} B\right)}_{= I_n - \frac{1}{n} B} H e = w^T w$$

$$\text{kde } w = \left(I_n - \tfrac{1}{n} B\right) H e \equiv K e$$

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

stačí tedy ukázat, že w a $\hat{e}$ jsou nezávislé vektory

$$\begin{pmatrix} w \\ \hat{e} \end{pmatrix} = \begin{pmatrix} K \\ L \end{pmatrix} e$$ tzn. má vícerozměrné normální rozdělení

$$\operatorname{Cov}\begin{pmatrix} w \\ \hat{e} \end{pmatrix} = \begin{pmatrix} K \\ L \end{pmatrix} \operatorname{Cov} e \,(K^T, L^T) = \sigma^2 \begin{pmatrix} K K^T & \overbrace{K L^T}^{=0\,??} \\ L K^T & L L^T \end{pmatrix}$$

pokud ano $\Rightarrow$ $w, \hat{e}$ nezávislé

$$K L^T = \left(I_n - \tfrac{1}{n} B\right) \underbrace{H (I_n - H)}_{= H - H^2 = 0} = 0$$

Q.E.D.

TEST: zamítnout H_0: $\beta_1 = \beta_2 = \dots = \beta_m = 0$ pokud $F > F_{1-\alpha}(m, n-m-1)$

POZN: odvozeno pro $e_i \sim N(0, \sigma^2)$, obecně se používá i když to neplatí, pro velká n může být často zdůvodněno pomocí CLV

Tabulka ANOVA

Source	df	SS	MS	F
Regression	m	SSR	$MSR = \frac{SSR}{m}$	$\frac{MSR}{MSE}$
Residual	$n-(m+1)$	SSE	$MSE = \frac{SSE}{n-m-1} = s_n^2$	
Total	$n-1$	SST		
		R^2	$\bar{R}^2$	

Koeficient (vícenásobné) determinace R^2

podobně jako u jednorozměrné regrese, lze F-test chápat jako test významnosti R^2, ~~protože~~ definovaného jako

$$R^2 \equiv \frac{SSR}{SST} = 1 - \frac{SSE}{SST} \qquad (SST = SSR + SSE)$$

protože

$$F = \frac{\frac{SSR}{m}}{\frac{SSE}{n-m-1}} = \frac{n-m-1}{m}\left(\frac{\frac{SSR}{SST}}{\frac{SSE}{SST}}\right) = \frac{n-m-1}{m}\,\frac{R^2}{1-R^2}$$

což je ~~nelineární~~ rostoucí funkce R^2. (opět $R^2 \in \langle 0,1\rangle$)

POZN: R^2 je možné zvětšovat přidáváním ~~nových~~ nových proměnných x, i když jsou statisticky nevýznamné. (pro n LN proměnných x a n pozorování dostaneme „perfect fit“.) – overfitting vysvětlení:

$R^2 = 1 - \frac{SSE}{SST}$, kde SST je pevně dáno daty y

ale SSE může být sníženo přidáním proměnných x

minimalizujeme totiž $(y - X\beta)^T (y - X\beta)$

přes větší množinu β

tzn. $\frac{SSE}{SST}$ je nerostoucí funkce počtu proměnných a tedy

R^2 je neklesající funkce počtu proměnných

proto se někdy definuje upravený koeficient determinace

(adjusted coefficient of determination)

$$\bar{R}^2 = R^2_{adj} = 1 - \frac{\frac{SSE}{n-m-1}}{\frac{SST}{n-1}} = 1 - \frac{n-1}{n-m-1}\,\frac{SSE}{SST}$$

(s rostoucím m klesá SSE ale i $n-m-1$)

jistá kompenzace