01RAD - přednáška 6, 21.10.2025

Věta 3.5

Nechť $\boldsymbol{Y} = \boldsymbol{X}\beta + \boldsymbol{e}$ je LM $(**)$, kde $h(\boldsymbol{X}) = m+1$ a $\boldsymbol{e} \sim N_n(\boldsymbol{0}, \sigma^2 \boldsymbol{I}_n)$. Potom

1) $\widehat{\beta}$ a s_n^2 jsou nezávislé náhodné veličiny,
2) $(n-m-1)\dfrac{s_n^2}{\sigma^2} \sim \chi^2(n-m-1)$,
3) jestliže $v_i = (\boldsymbol{X}^T\boldsymbol{X})^{-1}_{ii}$, potom $T_i = \dfrac{\widehat{\beta}_i - \beta_i}{s_n\sqrt{v_i}} \sim t(n-m-1)$.
4) Nechť $\boldsymbol{C} \in \mathbb{R}^{r,m+1}$ takové, že $h(\boldsymbol{C}) = r$. Potom kvadratická forma

$$\frac{q}{\sigma^2} = \frac{(\widehat{\beta}-\beta)^T \boldsymbol{C}^T \left[\boldsymbol{C}(\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{C}^T\right]^{-1} \boldsymbol{C}(\widehat{\beta}-\beta)}{\sigma^2} \sim \chi^2(r).$$

Důkaz.

Výsledky z LA:

- Spektrální rozklad matice:
  $\boldsymbol{A}_{n\times n}$ symetrická matice $\Rightarrow$ existuje ortogonální matice $\boldsymbol{Q}$ a diagonální matice $\boldsymbol{\Lambda}$ tak, že $\boldsymbol{A} = \boldsymbol{Q}\boldsymbol{\Lambda}\boldsymbol{Q}^T$, sloupce $\boldsymbol{Q}$ jsou ON vlastní vektory matice $\boldsymbol{A}$ a diagonální prvky matice $\boldsymbol{\Lambda}$ jsou jim odpovídající vlastní čísla.
- $\boldsymbol{A}_{n\times n}$ idempotentní matice $\Rightarrow$ vlastní čísla jsou pouze 0 nebo 1 a $\mathrm{h}(\boldsymbol{A}) = \mathrm{tr}(\boldsymbol{A})$

Vlastnosti vektoru reziduí $\widehat{\boldsymbol{e}}$

Věta 3.6

Uvažujme model $\boldsymbol{Y} = \boldsymbol{X}\beta + \boldsymbol{e}$, kde $e_1, \ldots, e_n$ jsou nekorelované a $e_i \sim (0, \sigma^2)$. Nechť $\widehat{\beta}$ je OLS β a $\widehat{\boldsymbol{e}} = \boldsymbol{Y} - \widehat{\boldsymbol{Y}}$ je vektor reziduí. Potom platí:

1) $\mathrm{E}\,\widehat{\boldsymbol{e}} = 0, \quad \mathrm{Cov}(\widehat{\boldsymbol{e}}) = \sigma^2(\boldsymbol{I}_n - \boldsymbol{H})$,
2) pokud navíc $\boldsymbol{e} \sim N_n(0, \sigma^2 \boldsymbol{I}_n)$, potom $\widehat{\boldsymbol{e}} \sim N_n(\mathbf{0}, \sigma^2(\boldsymbol{I}_n - \boldsymbol{H}))$,
3) jestliže má model intercept, tj. $\beta_0 \neq 0$, potom $\sum_{i=1}^n \widehat{e}_i = 0$,
4) $\sum_{i=1}^n \widehat{e}_i \widehat{y}_i = 0$.

Důsledek: Použitím bodů 3) a 4) dostaneme (stejně jako u jednorozměrné regrese)

$$\sum_{i=1}^n (y_i - \overline{y})^2 = \sum_{i=1}^n (\widehat{y}_i - \overline{y})^2 + \sum_{i=1}^n (y_i - \widehat{y}_i)^2$$

tedy

$$SST = SSR + SSE \qquad \text{(v modelu s interceptem)}$$

Důkaz.

3.2 Gauss-Markov theorem

- e_i i.i.d $N(0, \sigma^2)$ $\Rightarrow$ OLS $\widehat{\boldsymbol{\beta}}$ je MLE, tzn. je eficientní (MVUE par. $\boldsymbol{\beta}$)
- nenormální chyby:
  - ukážeme, že OLS $\widehat{\boldsymbol{\beta}}$ je BLUE (best linear unbiased estimator) par. $\boldsymbol{\beta}$ (za jistých podmínek)
  - mohou ale existovat lepší lineární vychýlené odhady nebo nelineární odhady

Definice 3.1

Nechť $\boldsymbol{\beta}$ je vektor regresních parametrů v LM. Řekneme, že $\widehat{\boldsymbol{\beta}}$ je lineární odhad $\boldsymbol{\beta}$, jestliže každé $\widehat{\beta}_j$ je LK pozorování Y_i, $i = 1, \dots, n$, tedy

$$\widehat{\beta}_j = \sum_{i=1}^{n} a_{ij} Y_i, \quad j = 0, \dots, m\,.$$

V maticovém zápisu $\widehat{\boldsymbol{\beta}} = \boldsymbol{AY}$, kde $\boldsymbol{A}^T = (a_{ij})$, $i = 1, \dots, n$, $j = 0, \dots, m$.

Poznámka: Pokud v modelu $(**)$ platí $h(\boldsymbol{X}) = m + 1$, potom je OLS $\widehat{\boldsymbol{\beta}}$ lineární, neboť $\widehat{\boldsymbol{\beta}} = (\boldsymbol{X}^T \boldsymbol{X})^{-1} \boldsymbol{X}^T \boldsymbol{Y}$.

VĚTA 3.7 (Gauss-Markov)

Uvažujme model $\boldsymbol{Y} = \boldsymbol{X}\beta + \boldsymbol{e}$, kde matice $\boldsymbol{X}$ má plnou hodnost, $e_1, \dots, e_n$ jsou nekorelované a $e_i \sim (0, \sigma^2)$. Potom OLS odhad $\widehat{\beta}$ je BLUE parametru β.

Důkaz.

3.3 Testování modelu - tabulka ANOVA

Celkový F-test (overall F-test)

- Je model statisticky signifikantní? Tj. je alespoň jeden z koeficientů $\beta_1, \ldots, \beta_m$ nenulový?
- Mohli bychom testovat jednotlivé koeficienty $\mathrm{H}_0 : \beta_j = 0$ 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
- Navíc je zde problém multikolinearity, jejíž jedním efektem jsou velké stand. chyby odhadů. To může vést k akceptování všech koeficientů jeho 0, i když je model celkově významný.

Bylo by dobré mít jednu statistiku pro test

$$\mathrm{H}_0 : \beta_1 = \beta_2 = \cdots = \beta_m = 0 \quad \times \quad \mathrm{H}_1 : \exists i \in \widehat{m}\,, \beta_i \neq 0$$

ANOVA přístup pro jedn. regresi naznačuje, že statistika

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

by mohla být užitečná (vyplyne i z obecnějších přístupů k testování později)

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

$$\overline{\boldsymbol{X}} = \begin{pmatrix} \overline{x}_0 & \overline{x}_1 & \cdots & \overline{x}_m \\ \vdots & \vdots & & \vdots \\ \overline{x}_0 & \overline{x}_1 & \cdots & \overline{x}_m \end{pmatrix}_{n\times m+1} \qquad (\boldsymbol{X}_c)_{ij} = x_{ij} - \overline{x}_j, \quad i = 1,\dots,n,\ j = 1,\dots,m$$

Věta 3.8

V modelu $\boldsymbol{Y} = \boldsymbol{X}\beta + \boldsymbol{e}$ tvaru $(**)$, kde $h(\boldsymbol{X}) = m+1$, e_i jsou nekorelované a $e_i \sim (0,\sigma^2)$, $i = 1,\dots,n$, platí

$$\mathrm{E}\left[\frac{SSR}{m}\right] = \sigma^2 + \frac{\boldsymbol{\beta}^T(\boldsymbol{X}-\overline{\boldsymbol{X}})^T(\boldsymbol{X}-\overline{\boldsymbol{X}})\beta}{m} = \sigma^2 + \frac{\boldsymbol{\beta}_s^T\boldsymbol{X}_c^T\boldsymbol{X}_c\boldsymbol{\beta}_s}{m},$$

kde $\boldsymbol{\beta}_s = (\beta_1,\dots,\beta_m)^T$.

Věta z PRA: Nechť $Z = \boldsymbol{Y}^T\boldsymbol{A}\boldsymbol{Y}$ je kvadratická forma a nechť $\mathrm{E}\boldsymbol{Y} = \boldsymbol{\mu}$ a $\mathrm{Cov}\boldsymbol{Y} = \boldsymbol{\Sigma}$. Potom platí:

$$\mathrm{E}Z = \mathrm{tr}(\boldsymbol{A}\boldsymbol{\Sigma}) + \boldsymbol{\mu}^T\boldsymbol{A}\boldsymbol{\mu}.$$

Důkaz.

Poznámka 3.5

- pokud $\boldsymbol{\beta}_s = 0$, potom $\mathrm{E}\left(\frac{SSR}{m}\right) = \sigma^2 = \mathrm{E}s_n^2$, $\boldsymbol{\beta}_s \neq 0$ implikuje, že $\mathrm{E}\left(\frac{SSR}{m}\right) > \sigma^2$
- tedy velké hodnoty $F = \frac{SSR/m}{s_n^2}$ budou znamenat zamítnutí $H_0 : \boldsymbol{\beta}_s = 0$