

## Vliv změn měřítkování proměnných na hodnoty statistických charakteristik

Je zajisté užitečné vědět, které statistické charakteristiky modelu jsou dotčeny změnou měrových jednotek vysvětlujících veličin nebo vysvětlované veličiny. Pokud bychom toto neznali, byli bychom v pochybnostech, zda získané výsledky jsou nebo nejsou ovlivněny takovýmito změnami a nemohli bychom vyvozovat příslušné statistické závěry s náležitou jistotou.

Předpokládejme, že původní hodnoty vysvětlujících veličin jsou ve vektorech  $x_1, x_2, x_3, \dots, x_K$  a původní hodnoty vysvětlované proměnné jsou ve vektoru  $y$ .

Nechť nyní dojde ke změně měřítka, takže nové „měřítkové“ proměnné jsou

$\tilde{x}_1, \tilde{x}_2, \tilde{x}_3, \dots, \tilde{x}_K$ , resp.  $\tilde{y}$  a pro vztahy mezi původními a novými proměnnými platí vztahy

$$\tilde{x}_1 = c_1 \cdot x_1, \tilde{x}_2 = c_2 \cdot x_2, \tilde{x}_3 = c_3 \cdot x_3, \dots, \tilde{x}_K = c_K \cdot x_K, \quad \text{resp.} \quad \tilde{y}_i = c \cdot y_i, \quad \text{kde}$$

nenulové reálné koeficienty  $c_1, c_2, c_3, \dots, c_K$ , resp. reálný skalár  $c$  představují změny měřítek, tj. přepočtené hodnoty nových měrových jednotek vzhledem k jednotkám původním.

Abychom vyšetřili dopad změn měřítek na hodnoty statistických charakteristik, musíme vyjádřit vektorově/maticové vztahy mezi původními a změněnými modelovými proměnnými. Toho dosáhneme zápisem vztahů mezi *prostými* a *ovlnkovanými* veličinami.

$$(1) \quad \tilde{X} = X \cdot C \quad \text{resp.} \quad \tilde{y} = y \cdot c,$$

kde  $C$  je diagonální matice typu  $[k, k]$  s jednotlivými  $c_k$  na hlavní diagonále (ostatní prvky  $C$  jsou nulové) a  $c$  je skalární konstanta, kterou násobíme jednotlivé složky vektoru  $y$ . (Pokud každou složku vektoru  $y$  zestonásobíme, pak  $c = 100$ .)

Nyní zapíšeme vztahy mezi původními měrovými jednotkami, ve kterých jsou vyjádřeny jednotlivé proměnné, a změněnými měrovými jednotkami<sup>1</sup>.

V důsledku toho, že matice  $C$  je (jako diagonální matice s nenulovými prvky) regulární, můžeme psát

$$(2) \quad X = \tilde{X} \cdot C^{-1} \quad \text{resp.} \quad y = \tilde{y} \cdot c^{-1} = \frac{\tilde{y}}{c},$$

Nejprve vyšetříme jednotlivé fragmenty vystupující ve výrazech pro důležité statistické charakteristiky. Postupně máme<sup>2</sup>

$$(X'X)^{-1} = [(\tilde{X}C^{-1})' \tilde{X}C^{-1}]^{-1} = [C^{-1} \tilde{X}' \tilde{X} C^{-1}]^{-1} = C(\tilde{X}' \tilde{X})^{-1} C$$

$$X'y = (\tilde{X}C^{-1})' \tilde{y} c^{-1} = C^{-1} c^{-1} \tilde{X}' \tilde{y} \quad \text{a odtud dále}$$

$$a) \quad \hat{\beta} = (X'X)^{-1} X'y = C(\tilde{X}' \tilde{X})^{-1} C \cdot C^{-1} \tilde{X}' \tilde{y} \cdot c^{-1} = C(\tilde{X}' \tilde{X})^{-1} \tilde{X}' \tilde{y} \cdot c^{-1} = c^{-1} C \hat{\tilde{\beta}}.$$

<sup>1</sup> S ohledem na symetrii matice  $C$  platí:  $C = C'$ ,  $C^{-1} = C^{-1}$ '

<sup>2</sup> Ovlnkované symboly píšeme ve vztazích napravo; odpovídá to interpretaci, kdy se ptáme, jak budou *původní* veličiny touto změnou ovlivněny.

U dalších veličin dostáváme

$$\mathbf{b1)} \quad \hat{\beta}' X' y = \left[ c^{-1} C \hat{\beta} \right]' \cdot \left[ C^{-1} \tilde{X}' \tilde{y} c^{-1} \right] = c^{-2} \hat{\beta}' C \cdot C^{-1} \tilde{X}' \tilde{y} = c^{-2} \hat{\beta}' \tilde{X}' \tilde{y}.$$

$$\mathbf{b2)} \quad y' y = \left[ c^{-1} \tilde{y} \right] \left[ c^{-1} \tilde{y} \right] = c^{-2} \tilde{y}' \tilde{y}. \quad \mathbf{a\ odtud} \quad R^2 = \frac{\hat{\beta}' X' y}{y' y} = \frac{c^{-2} \hat{\beta}' \tilde{X}' \tilde{y}}{c^{-2} \tilde{y}' \tilde{y}} = \tilde{R}^2$$

$$\mathbf{c1)} \quad SST = \sum (y_t - \bar{y})^2 = \sum (\tilde{y}_t c^{-1} - \bar{y} c^{-1})^2 = c^{-2} \sum (\tilde{y}_t - \bar{y})^2 = c^{-2} \tilde{S} \tilde{S}' T$$

$$\mathbf{c2)} \quad SSE = \sum e_t^2 = \sum (y_t - X_t \hat{\beta})^2 = \sum (c^{-1} \tilde{y}_t - \tilde{X}_t C^{-1} c^{-1} C \hat{\beta})^2 = \\ = \sum (c^{-1} (\tilde{y}_t - \tilde{X}_t \tilde{\beta}))^2 = c^{-2} \tilde{S} \tilde{S}' E$$

Odtud dostaneme jednak opět např.

$$R^2 = 1 - \frac{SSE}{SST} = 1 - \frac{c^{-2} \tilde{S} \tilde{S}' E}{c^{-2} \tilde{S} \tilde{S}' T} = 1 - \frac{\tilde{S} \tilde{S}' E}{\tilde{S} \tilde{S}' T} = \tilde{R}^2, \quad \mathbf{jednak\ také}$$

$$s_e = \sqrt{\frac{SSE}{T-k}} = \sqrt{\frac{c^{-2} \tilde{S} \tilde{S}' E}{T-k}} = c^{-1} \tilde{s}_e \quad \mathbf{a\ tedy\ následně}$$

$$\mathbf{c3)} \quad S_{bb} = s_e^2 (X' X)^{-1} = c^{-2} \tilde{s}_e^2 C (\tilde{X}' \tilde{X})^{-1} C = c^{-2} C S_{\tilde{b}\tilde{b}} C$$

**d) t-statistiky měřítkovaných proměnných nedoznají žádných změn, neboť**

$$t_{b_k} = \frac{b_k}{s_{b_k}} = \frac{c^{-1} c_k \tilde{b}_k}{c^{-1} c_k \tilde{s}_{b_k}} = \frac{\tilde{b}_k}{\tilde{s}_{b_k}} = \tilde{t}_{b_k}$$

Zde jsme uplatnili okolnost, že matice  $C$  je diagonální. Máme tedy např.

$$b_k = c^{-1} \cdot c_k \tilde{b}_k$$

a vezmeme-li druhé odmocniny z diagonálních prvků  $S_{bb}$ , dostaneme podobně

$$s_{b_k} = c^{-1} c_k \cdot \tilde{s}_{b_k}$$

**e) Durbin-Watsonův koeficient autokorelace reziduí<sup>3</sup>, který je definován jako**

$$DW = \frac{\sum_{t=2}^T (e_t - e_{t-1})^2}{\sum_{t=1}^T e_t^2} = \frac{\sum_{t=2}^T (c^{-1} \tilde{e}_t - c^{-1} \tilde{e}_{t-1})^2}{\sum_{t=1}^T (c^{-1} \tilde{e}_t)^2} = \frac{c^{-2} \sum_{t=2}^T (\tilde{e}_t - \tilde{e}_{t-1})^2}{c^{-2} \sum_{t=1}^T \tilde{e}_t^2},$$

se rovněž nezmění. Rezidua se změni srovnatelně s vysvětlovanou proměnnou

$$e_t = y_t - \hat{y}_t = y_t - x_t \hat{\beta} = \tilde{y}_t c^{-1} - (\tilde{x}_t C^{-1}) \cdot (c^{-1} C \tilde{y}_t) = c^{-1} (\tilde{y}_t - \tilde{y}_t) = c^{-1} \tilde{e}_t \quad \mathbf{4}$$

<sup>3</sup> Znamená to mj., že žádná z obvyklých testových statistik nebude změnou měrových jednotek dotčena.

<sup>4</sup> Jako  $x_t$  jsme označili t-tý řádek matice  $X$

### Příklad:

Nejčastěji vyjadřujeme změnu měřítek v celých mocninách 10 jako měřítka měnících se součinitelů. Změna jednotek se zde provede pouhým posunem desetinné čárky. Přitom se hodnoty regresních koeficientů a k nim příslušných směrodatných odchylek změní přímo úměrně příslušnému  $10^k$ -násobku vysvětlované proměnné a současně nepřímo úměrně  $10^l$ -násobku vysvětlující veličiny, u které tento koeficient (resp. jeho směrodatná odchylka) stojí.

### Ilustrace:

Mějme regresi se dvěma vysvětlujícími proměnnými  $y_t = \alpha + \beta_1 x_{t2} + \beta_2 x_{t3}$

Předpokládejme, že měrové jednotky, v nichž vyjadřujeme vysvětlující proměnnou  $x_2$  ztisícínásobíme a jednotky, v nichž vyjadřujeme vysvětlující proměnnou  $x_3$ , zdesetinásobíme (ke změně úrovně konstanty  $x_1$  přímý důvod nemáme). Současně zestonásobíme hodnoty vysvětlované proměnné  $y_t$ . Budeme mít tedy modifikace  $\tilde{y}_t = 100 \cdot y_t$ ,  $\tilde{x}_{t1} = x_{t1}$ ,  $\tilde{x}_{t2} = 1000 \cdot x_{t2}$ ,  $\tilde{x}_{t3} = 10 \cdot x_{t3}$

Za uvedených předpokladů máme

$$C = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1000 & 0 \\ 0 & 0 & 10 \end{pmatrix}, \quad c = 100$$

Nejprve vyšetříme jednotlivé fragmenty vystupující ve výrazech pro důležité statistické charakteristiky. Postupně máme

$$\hat{\beta} = \frac{C}{c} \cdot \hat{\tilde{\beta}} = \begin{pmatrix} 1/100 & 0 & 0 \\ 0 & 1000/100 & 0 \\ 0 & 0 & 10/100 \end{pmatrix} \hat{\tilde{\beta}} = \begin{pmatrix} 0,01 & 0 & 0 \\ 0 & 10 & 0 \\ 0 & 0 & 0,1 \end{pmatrix} \hat{\tilde{\beta}}$$

Uvedené změny budou mít za následek to, že v nově určeném vektoru regresních koeficientů  $\tilde{\beta}$  bude úrovně konstanta 100-násobkem původní, regresní koeficient  $\tilde{\beta}_2$  u proměnné  $x_2$  desetinou a regresní koeficient  $\tilde{\beta}_3$  u proměnné  $x_3$  desetinásobkem původních regresních koeficientů  $\beta_2, \beta_3$ .

Na základě dříve dosažených výsledků lze tedy konstatovat, že

- regresní koeficienty se změní v proporcích 100 resp. 1/10 a 10 vůči původním
- směrodatné odchylky regresních koeficientů se rovněž změní v proporcích 100 resp. 1/10 a 10 vůči původním.
- t-statistiky regresních koeficientů se nezmění
- rozptyl závisle proměnné se zvýší  $10^4$  násobně
- reziduální směrodatná odchylka se zvýší 100 násobně
- koeficient determinace se nezmění
- Durbin-Watsonův koeficient autokorelace reziduí se nezmění

**Ještě jednou shrneme důsledky pro nejčastěji vyskytující se případy:**

**1. regresní koeficienty**

**1A. při změně měřítka závisle proměnné**  $\tilde{y}_t = d \cdot y_t$  ( $d$  skalár)  
změní se ve stejném poměru než daná závisle proměnná, protože

$$\hat{\tilde{\beta}} = (X' X)^{-1} X' \tilde{y} = (X' X)^{-1} X' d \cdot y = d \cdot (X' X)^{-1} X' y = d \cdot \hat{\beta}$$

**1B. při změně měřítka jediné (i-té) nezávisle proměnné**  $\tilde{x}_{ti} = d_i \cdot x_{ti}$   
( $d_i$  skalár nacházející se na i-tém místě hlavní diagonály matice D, ostatní diagonální prvky jsou 1) *změní se v opačném poměru než daná závisle proměnná*

$$\hat{\tilde{\beta}} = (\tilde{X}' \tilde{X})^{-1} \tilde{X}' y = (D' X' X D)^{-1} X D y = d_i^{-2} \cdot (X' X)^{-1} d_i X y = \frac{1}{d_i} \hat{\beta}$$

**1C. při změně měřítka závisle a jediné nezávisle proměnné**  $\tilde{y}_t = d \cdot y_t, \tilde{x}_{ti} = d_i \cdot x_{ti}$ ,

$$\hat{\tilde{\beta}} = (\tilde{X}' \tilde{X})^{-1} \tilde{X}' \tilde{y} = (D \cdot X' X D)^{-1} X \cdot D y = d_i^{-2} \cdot (X' X)^{-1} d_i X d \cdot y = \frac{d}{d_i} \hat{\beta}$$

**2. směrodatné odchylky parametrů**

**2A. při změně měřítka závisle proměnné**  $\tilde{y}_{ti} = d \cdot y_{ti}$  ( $d$  skalár)  
změní se ve stejném poměru jako závisle proměnná, protože

$$\tilde{s}_\beta = \sqrt{(\tilde{X}' \tilde{X})^{-1} \tilde{s}^2_\varepsilon} = \sqrt{(X' X)^{-1} (d \cdot s_\varepsilon)^2} = d \cdot \sqrt{(X' X)^{-1} s^2_\varepsilon} = d \cdot s_\beta$$

**2B. při změně měřítka nezávisle proměnné**  $\tilde{x}_{ti} = d_i \cdot x_{ti}$  ( $d_i$  skalár)  
změní se v opačném poměru než daná vysvětlovaná proměnná

$$\tilde{s}_\beta = \sqrt{(\tilde{X}' \tilde{X})^{-1} s^2_\varepsilon} = \sqrt{(D' X' X D)^{-1} s^2_\varepsilon} = \frac{1}{d_i} \sqrt{(X' X)^{-1} s^2_\varepsilon} = \frac{1}{d_i} \cdot s_\beta$$

**2C. při změně měřítka závisle i některé z nezávisle proměnných**

$$\tilde{s}_\beta = \sqrt{(\tilde{X}' \tilde{X})^{-1} \tilde{s}^2_\varepsilon} = \sqrt{(c^{-1} X' X c^{-1})^{-1} (d \cdot s_\varepsilon)^2} = \frac{d}{c} \sqrt{(X' X)^{-1} s^2_\varepsilon} = \frac{d}{c} \cdot s_\beta$$

**t-statistiky regresních parametrů**

**3A. při změně měřítka závisle proměnné**  $\tilde{y}_{ti} = d \cdot y_{ti}$  ( $d$  skalár)

$$\tilde{t}_\beta = \frac{\tilde{\beta}}{\tilde{s}_\beta} = \frac{d \cdot \beta}{d \cdot s_\beta} = t_\beta \quad \text{tj. nezmění se}$$

**3B. při změně měřítka nezávisle proměnné**

$$\tilde{t}_\beta = \frac{\tilde{\beta}}{\tilde{s}_\beta} = \frac{c^{-1} \cdot \beta}{c^{-1} \cdot s_\beta} = t_\beta \quad \text{tj. nezmění se}$$

**3C. při změně měřítka závisle i některé z nezávisle proměnných**

$$\tilde{t}_\beta = \frac{\tilde{\beta}}{s_\phi} = \frac{c^{-1}d \cdot \beta}{c^{-1}d \cdot s_\beta} = t_\beta \quad \text{tj, nezmění se}$$

#### **4 koeficient determinace**

**4A. při změně měřítka závisle proměnné**

**4B. při změně měřítka nezávisle proměnné**

**4C. při změně měřítka závisle i některé z nezávisle proměnných**

#### **5 standardní chyba (směrodatná odchylka) reziduí**

**5A. při změně měřítka závisle proměnné**

**5B. při změně měřítka nezávisle proměnné**

**5C. při změně měřítka závisle i některé z nezávisle proměnných**

#### **3. rozptyl závisle proměnné**

**6A. při změně měřítka závisle proměnné**

**6B. při změně měřítka nezávisle proměnné**

**6C. při změně měřítka závisle i některé z nezávisle proměnných**