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 $x_{.1}, x_{.2}, x_{.3}, \ldots, x_{.K}$ a původní hodnoty vysvětlované proměnné jsou ve vektoru y_t.

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

$\tilde{x}_{.1}, \tilde{x}_{.2}, \tilde{x}_{.3}, \ldots, \tilde{x}_{.K}$, resp. $\tilde{y}$ a pro vztahy mezi původními a novými proměnnými platí vztahy

$$\tilde{x}_{.1} = c_1.x_{.1}, \tilde{x}_{.2} = c_2.x_{.2}, \tilde{x}_{.3} = c_3.x_{.3}, \ldots, \tilde{x}_{.K} = c_K x_{.K} \quad , \quad \textbf{resp.} \quad \tilde{y}_t = c.y_t \;, \text{ kde}$$

nenulové reálné koeficienty $c_1, c_2, c_3, \ldots, c_K$, 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) $$\tilde{X} = X.C \qquad \textbf{resp.} \qquad \tilde{y} = y.c \quad ,$$

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[1].

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

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

Nejprve vyšetříme jednotlivé fragmenty vystupující ve výrazech pro důležité statistické charakteristiky. Postupně máme[2]

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

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

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

---

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

2 Ovlnkované symboly píšeme ve vztazích napravo; odpovídá to interpretaci, kdy se tážeme, jak budou *původní* veličiny touto změnou ovlivněny.

U dalších veličin dostáváme

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

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

c1) $SST = \sum (y_t - \bar{y})^2 = \sum (\tilde{y}_t c^{-1} - \bar{y} c^{-1})^2 = c^{-2} \sum (\tilde{y}_t - \bar{y})^2 = c^{-2} \widetilde{SST}$

c2)
$$SSE = \sum \varepsilon_t^{\,2} = \sum (y_t - X_t \hat{\beta})^2 = \sum (c^{-1} \tilde{y}_t - \tilde{X}_t C^{-1} c^{-1} C \tilde{\beta})^2 =$$
$$= \sum (c^{-1} (\tilde{y}_t - \tilde{X}_t \tilde{\beta}))^2 = c^{-2} \widetilde{SSE}$$

Odtud dostaneme jednak opět např.

$R^2 = 1 - \dfrac{SSE}{SST} = 1 - \dfrac{c^{-2} \widetilde{SSE}}{c^{-2} \widetilde{SST}} = 1 - \dfrac{\widetilde{SSE}}{\widetilde{SST}} = \tilde{R}^2$ **, jednak také**

$s_e = \sqrt{\dfrac{SSE}{T-k}} = \sqrt{\dfrac{c^{-2} \widetilde{SSE}}{T-k}} = c^{-1} \tilde{s}_e$ **a tedy následně**

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

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

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

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

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

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

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

e) Durbin-Watsonův koeficient autokorelace reziduí[3], který je definován jako

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

se rovněž nezmění. Rezidua se změní srovnatelně s vysvětlovanou proměnnou

$e_t = y_t - \hat{y}_t = y_t - x_{t.} \hat{\beta} = \tilde{y}_t c^{-1} - (\tilde{x}_{t.} C^{-1}) . (c^{-1} C \tilde{y}_t) = c^{-1} (\tilde{y}_t - \hat{\tilde{y}}) = c^{-1} \tilde{e}_t$ **[4]**

---

[3] Znamená to mj., že žádná z obvyklých testových statistik nebude změnou měrových jednotek dotčena.

[4] Jako $x_{.t}$ jsme označili t-tý řádek matice X

Příklad:

Nejčastěji vyjadřujeme změnu měřítek v celých mocninách 10 jako měřítko měnících součinitelů. Změna jednotek se provede pouhým posunem desetinné čárky. Tím se regresní koeficienty a jejich směrodatné odchylky převádějí do původních měrových jednotek toliko posunutím desetinné čárky.

Ilustrace:

Mějme regresi se dvěma vysvětlujícími proměnnými $y_t = \alpha + \beta_1 x_{t1} + \beta_2 x_{t2}$

Předpokládejme, že měrové jednotky, v nichž vyjadřujeme vysvětlující proměnnou x_1 **ztisícinásobíme a jednotky, v nichž vyjadřujeme vysvětlující proměnnou** x_2**, zdesetinásobíme** (*ke změně úrovňové konstanty* x_0 *přímý důvod nemáme*). **Současně zestonásobíme hodnoty vysvětlované proměnné** y_t**. Budeme mít tedy modifikace**

$\widetilde{y}_t = 100.y_t$, $\widetilde{x}_{t0} = x_{t0}$, $\widetilde{x}_{t1} = 1000.x_{t1}$, $\widetilde{x}_{t2} = 10.x_{t2}$

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

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

Změny budou mít za následek, že v nově určeném vektoru regresních koeficientů $\widetilde{\beta}$ **bude úrovňová konstanta 100-násobkem původní, regresní koeficient** $\widetilde{\beta}_1$ **u proměnné** $x_{.1}$ **desetinásobkem a regresní koeficient** $\widetilde{\beta}_2$ **u proměnné** $x_{.2}$ **desetinou původních regresních koeficientů** β_1, β_2.

Na základě dříve dosažených výsledků lze tedy konstatovat, že

a) regresní koeficienty ***se změní v proporcích 100 resp. 10 a 1/10 vůči původním***

b) směrodatné odchylky regresních koeficientů ***se*** *rovněž* ***změní v proporcích 100 resp. 10 a 1/10 vůči původním .***

c) t-statistiky regresních koeficientů ***se nezmění***

c) rozptyl závisle proměnné ***se zvýší*** $\boldsymbol{10^4}$ ***násobně***

e) reziduální směrodatná odchylka ***se zvýší 100 násobně***

f) koeficient determinace ***se nezmění***

g) 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 $\widetilde{y}_t = d.y_t$ (d skalár)

změní se ve stejném poměru než daná závisle proměnná, protože

$$\hat{\widetilde{\beta}} = (X'X)^{-1} X \, \widetilde{y} = (X'X)^{-1} X \, d.y = d.(X'X)^{-1} X \, y = d.\beta$$

1B. při změně měřítka jediné (i-té) **nezávisle proměnné** $\widetilde{x}_{ti} = d_i.x_{ti}$

(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*

$$\hat{\widetilde{\beta}} = (\widetilde{X}'\widetilde{X})^{-1}\widetilde{X}\, y = (D'X'XD)^{-1} XD\, y = d_i^{-2}.(X'X)^{-1} d_i X\, y = \frac{1}{d_i}.\hat{\beta}$$

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

$$\hat{\widetilde{\beta}} = (\widetilde{X}'\widetilde{X})^{-1}\widetilde{X}\,\widetilde{y} = (D.X'XD)^{-1} X.D\, dy = d_i^{-2}.(X'X)^{-1} d_i X\, d.y = \frac{d}{d_i}\hat{\beta}$$

2. ***<u>směrodatné odchylky parametrů</u>***

2A při změně měřítka závisle proměnné $\widetilde{y}_{ti} = d.y_{ti}$ (d skalár)

změní se ve stejném poměru jako závisle proměnná, protože

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

2B při změně měřítka nezávisle proměnné $\widetilde{x}_{ti} = d_i.x_{ti}$ (d_i skalár)

změní se v opačném poměru než daná vysvětlovaná proměnná

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

2C při změně měřítka závisle i některé z nezávisle proměnných

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

<u>t-statistiky regresních parametrů</u>

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

$$\widetilde{t}_\beta = \frac{\widetilde{\beta}}{\widetilde{s}_\phi} = \frac{d.\beta}{d.s_\beta} = t_\beta$$

tj.nezmění se

3B. při změně měřítka nezávisle proměnné

$$\widetilde{t}_\beta = \frac{\widetilde{\beta}}{\widetilde{s}_\phi} = \frac{c^{-1}.\beta}{c^{-1}.s_\beta} = t_\beta$$

tj.nezmění se

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

$$\tilde{t}_\beta = \frac{\tilde{\beta}}{s_\phi} = \frac{c^{-1}d.\beta}{c^{-1}d.s_\beta} = t_\beta$$ *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