Koeficient determinace

základní definice jak pro centrované, tak pro necentrované veličiny

$$R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST}, \text{ kde}$$

SSR je regresní součet čtverců $SSR = \sum_{i=1}^{n} \left(\hat{y}_i - \bar{\hat{y}}\right)^2$

SST je celkový součet čtverců $SST = \sum_{i=1}^{n} \left(y_i - \bar{y}\right)^2$

SSE je reziduální součet čtverců $SSE = \sum_{i=1}^{n} \left(e_i\right)^2 = \sum_{i=1}^{n} \left(y_i - \hat{y}\right)^2$

přičemž $\bar{y} = \frac{\sum_{i=1}^{n} y_i}{n}$; $\bar{\hat{y}} = \frac{\left(\sum_{i=1}^{n} \hat{y}_i\right)}{n}$ a dále platí $\bar{y} = \bar{\hat{y}}$

Poznámka : u SSE není třeba odečítat nulový průměr, neboť víme, že $\sum_{i=1}^{n} e_i = 0$

Dále víme, že platí $\boldsymbol{y} = \boldsymbol{\beta} + \boldsymbol{\varepsilon}$, resp. $\hat{\boldsymbol{y}} = \boldsymbol{Xb} = \boldsymbol{X}\left(\boldsymbol{X'X}\right)^{-1} \boldsymbol{X'y}$ tj. $\hat{\boldsymbol{y}} = M \cdot \boldsymbol{y}$, kde $M = \boldsymbol{X}\left(\boldsymbol{X'X}\right)^{-1} \boldsymbol{X'}$

Dále uvažujme centrované proměnné $\boldsymbol{y}^*$ a $\boldsymbol{x}_1^*, \ldots, \boldsymbol{x}_k^*$, které jsou vyjádřeny jako odchylky od svých průměrů $\boldsymbol{x}y_i^* = y_i - \bar{y}$; $x_{ij}^* = x_{ij} - \bar{x}_j$ $j = 1, \ldots, k$:

Vyjádříme-li SSE výrazem $\left(\boldsymbol{y}^* - \boldsymbol{X}^*\boldsymbol{b}\right)' \left(\boldsymbol{y}^* - \boldsymbol{X}^*\boldsymbol{b}\right) = \boldsymbol{e'e}$, pak platí

$$\boldsymbol{R^2} = \boldsymbol{1} - \frac{\boldsymbol{SSE}}{\boldsymbol{SST}} = \frac{\boldsymbol{y}^{*\prime}\boldsymbol{y}^* - \left(\boldsymbol{y}^* - \boldsymbol{X}^*\boldsymbol{b}\right)' \left(\boldsymbol{y}^* - \boldsymbol{X}^*\boldsymbol{b}\right)}{\boldsymbol{y}^{*\prime}\boldsymbol{y}'} =$$

$$= \frac{\boldsymbol{y}^{*\prime}\boldsymbol{y}^* - \boldsymbol{y}^{*\prime}\boldsymbol{y}^* + \boldsymbol{b'X}^{*\prime}\boldsymbol{y}^*}{\boldsymbol{y}^{*\prime}\boldsymbol{y}^*} = \frac{\boldsymbol{b'X}^{*\prime}\boldsymbol{y}^*}{\boldsymbol{y}^{*\prime}\boldsymbol{y}^*} = \frac{\left(\boldsymbol{X}^*\boldsymbol{b}\right)'\boldsymbol{y}^*}{\boldsymbol{y}^{*\prime}\boldsymbol{y}^*}$$

protože (a) $\boldsymbol{e'e} = \left(\boldsymbol{y} - \boldsymbol{Xb}\right)' \left(\boldsymbol{y} - \boldsymbol{Xb}\right) = \boldsymbol{y'y} - \boldsymbol{b'X'y} - \boldsymbol{y'Xb} + \boldsymbol{b'X'Xb}$

a dále (b)

$$\boldsymbol{e'X} = \left(\boldsymbol{y} - \boldsymbol{Xb}\right)' \boldsymbol{X} = \boldsymbol{y'X} - \boldsymbol{b'X'X} = \boldsymbol{y'X} - \left[\left(\boldsymbol{X'X}\right)^{-1} \boldsymbol{X'y}\right]' \boldsymbol{X'X} = \boldsymbol{y'X} - \boldsymbol{y'X}\left(\boldsymbol{X'X}\right)^{-1} \boldsymbol{X'X} = 0$$

Poznámka : vztahy (a) a (b) platí jak pro centrované, tak pro necentrované veličiny ..

Ověření vztahu SST=SSE+SSR obecně pro necentrované veličiny

SST = $\sum_1^T (y_i - \bar{y})^2 = \sum_1^T y_i^2 - 2\sum_1^T y_i \cdot \bar{y} + T \cdot \bar{y}^2 = \sum_1^T y_i^2 - T\bar{y}^2$

SSE = $\sum_1^T (y_i - \hat{y}_i)^2 = \sum_1^T y_i^2 - 2\sum_1^T y_i\hat{y}_i + \sum_1^T \hat{y}_i^2$

SSR = $\sum_1^T \left(\hat{y}_i - \bar{\hat{y}}\right)^2 = \sum_1^T \hat{y}_i^2 - 2\sum_1^T \hat{y}_i \bar{\hat{y}} + T \cdot \bar{\hat{y}}^2$

SSE + SSR =

= $\sum_1^T (y_i - \hat{y}_i)^2 + \sum_1^T \left(\hat{y}_i - \bar{\hat{y}}^2\right) = \sum_1^T y_i^2 - 2\sum_1^T y_i\hat{y}_i + \sum_1^T \hat{y}_i^2 + \sum_1^T \hat{y}_i^2 - 2\sum_1^T \hat{y}_i \bar{\hat{y}} + T \cdot \bar{\hat{y}}^2 =$

$= \sum_1^T y_i^2 - T \cdot \bar{\hat{y}}^2$ (= SST), neboť

$-2\sum_1^T y_i\hat{y}_i + \sum_1^T \hat{y}_i^2 + \sum_1^T \hat{y}_i^2 = -2\sum_1^T y_i\hat{y}_i + 2\sum_1^T \hat{y}_i^2 = 0$, protože

$\sum_1^T y_i\hat{y}_i = \sum_1^T (\hat{y}_i + \hat{u}_i) \cdot \hat{y}_i = \sum_1^T \hat{y}_i^2 + \sum_1^T \hat{u}_i\hat{y}_i = \sum_1^T \hat{y}_i^2$ a dále

$$\bar{y} = \frac{\sum y_i}{T} = \frac{\sum (\hat{y}_i + \hat{u}_i)}{T} = \frac{\sum \hat{y}_i}{T} + \frac{\sum \hat{u}_i}{T} = \bar{\hat{y}}$$