# ***Stanovení chyby předpovědi***

**Uvažujme opět standardní normální lineární regresní model , u kterého jsme již předtím ověřili jeho způsobilost k prognózování minimálně ve smyslu ověření stability modelové struktury.**

**Budeme pracovat s modelem v zápisu *( pro predikované období o délce m )***

**(1)** $$y_p = X_p.\beta + \varepsilon_p$$ **, kde**

$y_p$ **je [*m-členný*] sloupcový vektor hodnot vysvětlované (endogenní) proměnné, kterých nabude v predikovaném období**

$X_p$ **je [*m x k– rozměrná*] matice hodnot vysvětlujících (exogenních) proměnných realizovaných v predikovaném období**

$\beta$ **je [ *k-členný* ] sloupcový vektor regresních parametrů ( hodnoty parametrů přejímáme beze změn z pozorovaného období)**

$\varepsilon_p$ **je [ *m-členný* ] sloupcový vektor náhodných složek predikovaného období**

**<u>Předpokládáme přitom, že</u>**

**a)** $E(\varepsilon) = 0_T$ **,** $E(\varepsilon_p) = 0_m$**, tzn. že náhodné složky s nulovou střední hodnotou v pozorovaném období zůstanou centrované i v předpovídaném období.**

**b)** $Var(\varepsilon_p^{(j)}) = Var(\varepsilon^{(i)}) = \sigma^2$**,** $i = 1,2,\dots,T; j = 1,2,\dots,m$ **; tzn. že rozptyly náhodných složek pozorovaného i predikovaného období jsou shodné ( rovné $\sigma^2$ ) .**

**c)** $E(\varepsilon_p^{(j)}.\varepsilon^{(i)}) = 0$ **,** $i = 1,2,\dots,T; j = 1,2,\dots,m$**; tzn. že náhodné složky predikovaného období jsou nekorelované navzájem i s náhodnými složkami pozorovaného období. Souhrnná kovarianční matice $\tilde{\Sigma}_{[T+m;T+m]}$ má tedy tvar**

$$\tilde{\Sigma} = Cov\begin{pmatrix} \varepsilon \\ \varepsilon_p \end{pmatrix} = \begin{pmatrix} \sigma^2 I_T & 0 \\ 0 & \sigma^2 I_m \end{pmatrix}$$

**d) Náhodné složky jsou nekorelované s vysvětlujícími proměnnými v pozorovaném i v predikovaném období, tj.** $E(X_j'\varepsilon) = 0$ **pro** $j = 1,2,\dots,T$
$E(X_p'\varepsilon_p) = 0$ **pro** $l = 1,2,\dots,m$ **.**

**Model můžeme přehledněji znázornit maticovým zápisem**

$$\begin{pmatrix} y_1 \\ y_2 \\ \dots\dots \\ y_T \\ y_{T+1} \\ \dots\dots \\ y_{T+m} \end{pmatrix} = \begin{pmatrix} x_{11} & x_{12} & x_{13} & \dots\dots & x_{1k} \\ x_{21} & x_{22} & x_{23} & \dots\dots & x_{2k} \\ \dots\dots & \dots\dots & \dots\dots & \dots\dots & \dots\dots \\ x_{T,1} & x_{T,2} & x_{T,3} & \dots\dots & x_{T,k} \\ x_{T+1,1} & x_{T+1,2} & x_{T+1,3} & \dots\dots & x_{T+1,k} \\ \dots\dots & \dots\dots & \dots\dots & \dots\dots & \dots\dots \\ x_{T+m,1} & x_{T+m,2} & x_{T+m,3} & \dots\dots & x_{T+m,k} \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \dots\dots \\ \beta_k \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \dots\dots \\ \varepsilon_T \\ \varepsilon_{T+1} \\ \dots\dots \\ \varepsilon_{T+m} \end{pmatrix}.$$

**Bodovou předpověď m-členného vektoru závisle proměnné získáme pomocí predikční funkce**[1] **tvaru**

**(2)** $$y_p^* = X_p.b = X_p.(X'X)^{-1}X'y$$ **, kde**

$X$ **je [T*x k– rozměrná*] matice pozorovaných hodnot k vysvětlujících proměnných**

**Predikční funkce (2), jež je zřejmě lineární funkcí $y$, je nestrannou funkcí vektoru předpovědí $y_p$ v tom smyslu, že pro ni platí**

**(3)** $$E(y_p^*) = E(X_p.b) = X_p E(b) = X_p.\beta = E(y_p)\ .$$

[ $X_p$ je nestochastická ] [ b je nestranný odhad $\beta$ ]

**To znamená, že střední hodnota predikční funkce je rovna střední hodnotě $y_p$.**[2]

***Poznámka*: Neznamená to však , že $E(y_p^*) = y_p$**[3]**.**

**Lze ukázat, že predikční funkce (2) má ze všech nestranných predikčních funkcí** (lineárními v $y$) **nejmenší** v maticovém smyslu **kovarianční matici vektoru reziduí v období předpovědi,** tj. následujícího vektoru $e_p$ **:**

**Vektor chyb $d_p$ definujeme jako odchylku** (chybu) **předpovídané hodnoty $y_p^*$ od skutečné $y_p$ :**

**(4)** $$d_p = y_p^* - y_p = X_p(b-\beta) - \varepsilon_p$$

**Všimněme si, že zdrojem chyb předpovědi vektoru skutečných hodnot $y_p$ je jednak variabilita náhodné složky $\varepsilon_p$, jednak výběrová chyba odhadové funkce $b$** *zpravidla pořízené metodou nejmenších čtverců OLS* **:**

**Střední chyba vektoru chyb predikčního vektoru $d_p$ v (4) je nulová, protože platí**

$$E(d_p) = E(y_p^* - y_p) = X_p E(b-\beta) - E\varepsilon_p = 0.$$

**Před odvozením kovarianční matice tohoto vektoru upravíme $d_p$ do tvaru**

**(5)** $$d_p = X_p.b - X_p\beta - \varepsilon_p = -\varepsilon_p - X_p\beta + X_p(X'X)^{-1}X'y =$$
$$= -\varepsilon_p - X_p\beta + X_p(X'X)^{-1}X'(X\beta + \varepsilon) =$$
$$= -\varepsilon_p - X_p\beta + X_p\beta + X_p(X'X)^{-1}X'\varepsilon = -\varepsilon_p + X_p(X'X)^{-1}X'\varepsilon$$

***Vektor reziduí předpovídaných hodnot závisí tedy na náhodných složkách pozorovaného i náhodných složkách predikovaného období.***

---

[1] Predikční funkce je náhodný vektor o m složkách.
[2] Jde o totožnost středních hodnot stejnolehlých složek příslušných náhodných vektorů.
[3] Podle (1) je $y_p = X_p.\beta + \varepsilon_p$ a náhodnost vektoru $y_p$ pramení z náhodnosti $\varepsilon_p$ , zatímco v případě $y^*_p$ vyplývá náhodnost tohoto vektoru z náhodnosti $b$ ( a tudíž z $\varepsilon$ ).

**Za přijatých předpokladů a) – d) lze pro <u>kovarianční matici</u>** náhodného **<u>predikčního vektoru</u> (5) psát :**

$$Cov(d_p) = E\left[-\varepsilon_p + X_p(X'X)^{-1}X'\varepsilon\right].\left[-\varepsilon_p + X_p(X'X)^{-1}X'\varepsilon\right]' =$$ [4]

[nekorelovanost ε, εp]

$$= E\left[-\varepsilon_p.\varepsilon_p'\right] + E\left[X_p(X'X)^{-1}X'\varepsilon\varepsilon'X(X'X)^{-1}X_p'\right] =$$

[ $E\,\varepsilon_p.\varepsilon_p' = \sigma^2 I_m$ ] [ $E\varepsilon.\varepsilon' = \sigma^2 I_T$ ]

$$= o^2.\left[I_m + X_p(X'X)^{-1}X_p'\right]$$

**<u>Poznámka :</u> speciální případ jediné predikované hodnoty tzn. případ m = 1**

$$Cov(d_p) = o^2.\left[1 + x_p(X'X)^{-1}x_p'\right] =$$

***V tomto případě jsou $y_p$ a $\varepsilon_p$ skalární hodnoty a matice $X_p$ se redukuje na řádkový vektor $x_p$ o $k$ složkách.***

**Standardní chybu (směrodatnou odchylku) bodové předpovědi skutečné hodnoty vysvětlované proměnné dostaneme nahrazením rozptylu $\sigma^2$ jeho nestranným odhadem $s^2$** *získaným například metodou OLS.* **Pak**

$$s_{d_p} =_{OLS} s.\sqrt{\left[1 + X_p(X'X)^{-1}X_p'\right]}$$ [5]

**Lze ukázat, že vyhovuje-li model předpokladu normality, pak bodová predikční funkce** založená na odhadu β metodou OLS **má též normální rozdělení, konkrétně :**

**(6)** $$y^*_p \approx N\left[X_p.\beta\,;\sigma^2 X_p(X'X)^{-1}X_p'\right],$$ protože zřejmě

**(3*)** $$E[y_p^*] = E[X_p b] = X_p b = E\,[y_p]$$ **.** a dále

$$Cov(y_p^*) = Cov(X_p.b) = X_p.Cov(b).X_p' = X_p.\sigma^2(X'X)^{-1}X_p' = \sigma^2 X_p.(X'X)^{-1}X_p'$$

[ $Cov(b) = \sigma^2.(X'X)^{-1}$ ] [ $\sigma^2$ je skalár]

---

[4] Bereme v úvahu nekorelovanost $\varepsilon$ a $\varepsilon_p$ .

[5] Pokud m=1, pak $X_p$ je vektor a veličina $x_p(X'X)^{-1}x_p'$ je skalár , pokud m >1, pak se vezme diagonální prvek matice $X_p(X'X)^{-1}X_p'$ pořadím příslušný dané složce vektoru parametrů β.