

Standardní Normální lineární regresní model

K dříve vysloveným předpokladům o veličinách standardního lineárního regresního modelu připojíme další předpoklad :

(e) T-rozměrný vektor náhodných složek má vícerozměrné normální rozdělení opět s nulovým vektorem středních hodnot a s diagonální kovarianční maticí , tzn.

$$\Sigma = \sigma^2 I_T, \text{ stručněji zapsáno } \varepsilon \approx N(0; \sigma^2 I_T)$$

Jak známo, sdružená hustota T-rozměrného normálního rozdělení má v případě nezávislých náhodných veličin tvar :

$$f(\varepsilon) = \frac{1}{(2\pi)^{T/2} \sigma^T} \exp\left(-\frac{\varepsilon' \varepsilon}{2\sigma^2}\right)$$

Věta 2 (pro standardní normální lineární regresní model)

Za podmínek věty 1 (Gauss-Markovovy) a dále za dodatečného předpokladu o T-rozměrném rozdělení vektoru náhodných složek $\varepsilon \approx N(0, \sigma^2 I_T)$ lze ukázat, že :

(2A) Odhadová funkce ${}_{OLS}b = (X'X)^{-1} X' y$ je také normálně rozdělena s vektorem středních hodnot (rovným skut. parametrům) β a s kovarianční maticí $\sigma^2 (X'X)^{-1}$.

(2B) Náhodná veličina $\frac{e'e}{\sigma^2}$ resp. (jinak zapsaná jako $\frac{(T-k)s^2}{\sigma^2}$) má χ^2 -rozdělení o $(T-k)$ stupních volnosti. Počet stupňů volnosti je určen rozdílem mezi počtem pozorování T a počtem vysvětlujících proměnných k .

(2C) Náhodné veličiny $\frac{(T-k)s^2}{\sigma^2}$ a ${}_{OLS}b - \beta$ jsou vzájemně nezávislé.

(2D) Odhady vektoru parametrů β získané metodou nejmenších čtverců ${}_{OLS}b$ a metodou maximální věrohodnosti ${}_{ML}b$ jsou identické, platí tedy ${}_{OLS}b = {}_{ML}b$

Výše uvedená tvrzení postupně dokážeme:

Důkaz tvrzení (2A) Odhadovou funkci ${}_{OLS}b$ lze zřejmě vyjádřit ve tvaru

$$(X'X)^{-1} X' y = (X'X)^{-1} X'(X\beta + \varepsilon) = \beta + (X'X)^{-1} X' \varepsilon, \text{ kde}$$

na pravé straně je nestochastické povahy pouze vektor náhodných složek ε . Odhadová funkce ${}_{OLS}b$ je tedy lineární kombinací složek náhodného vektoru ε , přičemž tato lineární kombinace má nestochastické prvky (a navíc obsahuje nestochastický vektor β). Jak známo ze statistické teorie, lineární kombinace složek náhodného vektoru s normálním rozdělením má též normální rozdělení. Zbývá určit střední hodnotu a kovarianční matici tohoto vektoru . Zřejmě platí :

$$\begin{aligned} E({}_{OLS}b) &= E(X'X)^{-1} X' y = E[\beta + (X'X)^{-1} X' \varepsilon] = E\beta + E(X'X)^{-1} X' \varepsilon \\ &= \beta + (X'X)^{-1} X' E\varepsilon = \beta \quad \text{neboť dle předpokladu (a) } E\varepsilon = 0. \quad \text{Podobně} \end{aligned}$$

$$\begin{aligned} \text{Cov}({}_{OLS}b) &= E \left[(b - \beta) \cdot (b - \beta)' \right] = E \left[(X'X)^{-1} X' \varepsilon \varepsilon' X (X'X)^{-1} \right] = (X'X)^{-1} X' E \varepsilon \varepsilon' X (X'X)^{-1} \\ &= (X'X)^{-1} X' \sigma_{\varepsilon}^2 I_T X \cdot (X'X)^{-1} = \sigma_{\varepsilon}^2 (X'X)^{-1}. \end{aligned}$$

Poznámka: Kovarianční matice $\text{Cov}({}_{OLS}b)$ není na rozdíl od kovarianční matice náhodných odchylek **diagonální, složky vektoru** ${}_{OLS}b$ **tedy zpravidla budou vzájemně zkorelovány. Rozptyl každé této složky je dán součinem** σ_{ε}^2 **a prvku ležícího na j-tém místě hlavní diagonály matice** $(X'X)^{-1}$ **- tento prvek označíme v dalším textu jako** V^{jj} .

Tvrzení (2B) Náhodná veličina $\frac{e'e}{\sigma^2}$ resp. (jinak zapsaná jako $\frac{(T-k)s^2}{\sigma^2}$) má χ^2 -rozdělení o $(T-k)$ stupních volnosti. Počet stupňů volnosti je dán rozdílem mezi počtem pozorování a počtem vysvětlujících proměnných.

Důkaz tvrzení (2B)

a) Víme, že $\varepsilon' \varepsilon$ představuje součet druhých mocnin T vzájemně nezávislých (tj. zde nekorelovaných a normálně rozdělených) náhodných veličin (jmenovitě náhodných složek regresní rovnice).

b) Dále víme, že analogický výraz $e'e$ pro rezidua lze vyjádřit zápisem

$$e'e = \varepsilon' M M \varepsilon, \text{ kde } M = I_T - X(X'X)^{-1} X', \text{ neboť } e = M \cdot \varepsilon$$

c) Matice M je symetrická a idempotentní, neboť pro ni platí

$$M M = M M = (I_T - X(X'X)^{-1} X') (I_T - X(X'X)^{-1} X') = I_T - X(X'X)^{-1} X' = M$$

Tedy $e'e$ je kvadratická forma s hodnotami danou hodnotami matice M . Hodnota idempotentní matice je rovna její stopě. Stopa matice M je přitom rovna $T-k$. Počet stupňů volnosti χ^2 -rozdělení veličiny $e'e$ je určen stopou matice M (je tedy roven $T-k$). Podle Cochranovy věty má výraz $\varepsilon' M \varepsilon / \sigma^2$ χ^2 -rozdělení s tolika stupni volnosti, jaká je hodnota matice M v kvadratické formě $\varepsilon' M \varepsilon$. (Rozptylem σ^2 dělíme proto, abychom získali součet T nezávislých náhodných veličin s rozdělením $N(0,1)$).

Zřejmě přitom platí

$$\text{Cov}(\varepsilon' M \varepsilon / \sigma^2) = I_T.$$

d) Konečně je snadné ukázat, že platí

$$\frac{e'e}{\sigma^2} = \frac{(T-k)s^2}{\sigma^2}$$

Výraz $e'e$ představuje součet čtverců reziduí (též značen SSE). s^2 je nestranný odhad rozptylu náhodných složek, jehož tvar je právě

$$\frac{e'e}{T-k} = s^2 \quad \square.$$

Tvrzení (2C) Náhodné veličiny $\frac{(T-k)s^2}{\sigma^2}$ a ${}_{OLS}b - \beta$ jsou vzájemně nezávislé.

Důkaz tvrzení (2C)

Součet čtverců náhodných odchylek $\varepsilon'\varepsilon$ rozložíme následovně :

$$\varepsilon'\varepsilon = \varepsilon' [I_T - X(X'X)^{-1} X']\varepsilon + \varepsilon' [X(X'X)^{-1} X']\varepsilon = \varepsilon' M\varepsilon + \varepsilon' N\varepsilon \quad ,$$

Poznámka: matice M i N jsou idempotentní, s hodnotami $h(M) = T - k$ a $h(N) = k$.

Podle **Cochranovy věty** mají náhodné veličiny – kvadratické formy $P(\varepsilon), Q(\varepsilon)$ obsahující matice M, N tato rozdělení:

- kvadratická forma $P(\varepsilon) = \varepsilon' M\varepsilon / \sigma^2$ má rozdělení χ^2 o $T-k$ stupních volnosti
- kvadratická forma $Q(\varepsilon) = \varepsilon' N\varepsilon / \sigma^2$ má rozdělení χ^2 o k stupních volnosti

Obě tyto náhodné veličiny jsou vzájemně nezávislé, neboť platí $M.N = 0$ (stochastická nezávislost je takto „posuzována“ ortogonalitou matic M, N)

Dále, výrazy $\varepsilon' M\varepsilon$ a $\varepsilon' N\varepsilon$ lze rozepsat následovně :

$$\varepsilon' M\varepsilon = \varepsilon' [I_T - X(X'X)^{-1} X']\varepsilon = \varepsilon' M M\varepsilon = e'e = (T-k)s^2 \quad , \text{ kde } s^2 = \sigma^2$$

$$\varepsilon' N\varepsilon = \varepsilon' [X(X'X)^{-1} X']\varepsilon = \varepsilon' [X(X'X)^{-1} X' X(X'X)^{-1} X']\varepsilon = \varepsilon' N' N\varepsilon = (b - \beta)' X' X (b - \beta) ,$$

$$\text{neboť } b = \beta + (X'X)^{-1} X' \varepsilon \quad , \quad \text{tzn. } b - \beta = (X'X)^{-1} X' \varepsilon$$

Dále, protože podle předpokladu je $X'X$ (jako momentová matice) **pozitivně definitní** (a symetrická) matice, existuje regulární matice P rozměrů $[k, k]$ taková, že platí $X'X = P.P'$. Pak lze psát

$$\varepsilon' N\varepsilon = (b - \beta)' X' X (b - \beta) = (b - \beta)' P P' (b - \beta) = [P'(b - \beta)]' [P'(b - \beta)] .$$

Odtud plyne, že vektor $P'(b - \beta)$ a tedy též vektor $(b - \beta)$ - **protože matice P je nestochastická** - nezávislý na skaláru $s^2(T-k)$ a též na s^2 (o P totiž předpokládáme, že je nestochastická matice) . □.

Tvrzení (2D) Odhady vektoru parametrů β pořizené metodou nejmenších čtverců a metodou maximální věrohodnosti jsou identické.

Důkaz tvrzení (2D): Již jsme ukázali, že odhad β pořizený metodou OLS má tvar

$${}_{OLS}b = (X'X)^{-1} X' y$$

Zbývá tedy ukázat, že odhad pořizený metodou maximální věrohodnosti má stejný tvar . Při tomto ověření vyjdeme ze sdružené hustoty vektoru náhodných složek, která má tvar

$$f(\varepsilon) = \frac{1}{(2\pi)^{T/2} \sigma^T} \exp\left(-\frac{\varepsilon'\varepsilon}{2\sigma^2}\right) \quad \text{pro } \varepsilon \approx N(0; \sigma^2 I_T)$$

Tato sdružená hustota je současně tzv. věrohodnostní funkcí, v jejímž zápisu se projevuje rozdíl v chápání pozic (odhadovaných) parametrů a pozorovaných veličin. Píšeme tedy

$$f(y, X; \beta, \sigma^2) = L(\beta, \sigma^2; y, X)$$

Zápis sdružené hustoty $f(y, X; \beta, \sigma^2)$ (kde se pozorované hodnoty y, X uvádí před středníkem, zatímco parametry β, σ^2 za ním) **pohlíží na tvar (zde normálního) rozdělení jako na rozdělení náhodného vektoru s pevně danými (známými) určenými parametry β a σ^2 . Na hodnoty y zde pohlížíme, jakoby byly „generovány mechanismem“, který se řídí rozdělením náhodných složek ε (při dané matici vysvětlujících proměnných X).**

V zápise věrohodnostní funkce $L(\beta, \sigma^2; y, X)$ (kde oba tyto parametry uvádíme v zápise před středníkem) **se naopak na hledané parametry pohlíží jako na neznámé, které odhadujeme ze známých pozorovaných veličin regresní rovnice** (těmi jsou vektor y a matice X).

Věrohodnostní funkci, jejíž maximum hledáme, zapíšeme ve tvaru, do něhož zahrneme pozorované veličiny (a přirozeně též vektor β a skalár σ^2):

$$L(\beta, \sigma^2; y, X) = \frac{1}{(2\pi)^{T/2} \sigma^T} \exp\left(-\frac{(y - X\beta)'(y - X\beta)}{2\sigma^2}\right)$$

Přímá maximalizace této (nelineární) věrohodnostní funkce je komplikovaná.

Proto ji ekvivalentně maximalizujeme v logaritmovaném tvaru (logaritmus je spojitá rostoucí funkce, takže poloha původního maxima – ze statistického hlediska jde o modus – se po této transformaci nezmění):

$$L^*(\beta, \sigma^2; y, X) = \ln L(\beta, \sigma^2; y, X) = -T/2 \ln(2\pi) - T/2 \ln \sigma^2 - \frac{1}{2\sigma^2} (y - X\beta)'(y - X\beta)$$

Úkolem je maximalizovat $L^*(\beta, \sigma^2; y, X)$ vzhledem k β a σ^2 . Rovnocenným cílem je minimalizace kladně vzatého výrazu $L^{}(\beta, \sigma^2; y, X)$ (neboť $\sigma^T = (\sigma^2)^{T/2}$)**

$$\text{Min}_{\sigma^2, \beta} [T/2 \cdot \ln(2\pi) + T/2 \cdot \ln \sigma^2 + \frac{1}{2\sigma^2} (y - X\beta)'(y - X\beta)]$$

Při této minimalizaci postupně dostáváme :

$$(A1) \quad \frac{\partial L^{**}}{\partial \beta} = \frac{1}{2\sigma^2} \cdot (-2X'y + 2X'X\beta) = 0$$

$$(B1) \quad \frac{\partial L^{**}}{\partial \sigma^2} = \frac{T}{2\sigma^2} - \frac{1}{2\sigma^4} (y - X\beta)'(y - X\beta) = 0$$

Poznámka : při derivování podle rozptylu σ^2 zacházíme s touto veličinou jako s jediným (nedělitelným) symbolem

Řešením vztahu (A1) zřejmě dostaneme vektor odhadnutých parametrů ve tvaru

$${}_{ML} b = (X'X)^{-1} X'y$$

Následně nyní dosazením tohoto odhadu za β do vztahu (B1) obdržíme ¹

$$(B1^*) \quad \frac{\partial L^{**}}{\partial \sigma^2} = \frac{T}{2\sigma^2} - \frac{1}{2\sigma^4} (y - X(X'X)^{-1}X'y)'(y - X(X'X)^{-1}X'y) = 0$$

Po vynásobení dvojnásobkem rozptylu ($2\sigma^2$) dostaneme zjednodušení:

$$T - \frac{1}{\sigma^2} e'e = 0, \text{ odkud plyne } e'e = \sigma^2 T \text{ a následně } {}_{ML}\hat{\sigma}^2 = \frac{e'e}{T}.$$

Odhad reziduálního rozptylu získaný metodou maximální věrohodnosti má tedy tvar

$${}_{ML}\hat{\sigma}^2 = \frac{e'e}{T} = \frac{SSE}{T}$$

Poznámka: Všimněme si, že k odhadu koeficientů ${}_{ML}\beta$ jsme nepotřebovali operovat se σ^2 , zatímco následný odhad ${}_{ML}\hat{\sigma}^2$ byl již vázán na předtím pořízený *odhad ${}_{ML}\beta$ (při jinak odhadnutém β bychom mohli získat obecně jiný odhad pro σ^2).

Měli bychom ještě ukázat, že získaný výraz pro ${}_{ML}\beta$ dává skutečně minimum (nikoliv maximum nebo sedlový bod). To dokážeme, pokud druhý diferenciál vztahu pro L^{**} vede k matici, která je pozitivně definitní. Skutečně lze snadno ověřit, že

$$(B2) \quad \frac{\partial^2 L^{**}}{\partial \beta^2} = \frac{X'X}{\sigma^2} > 0, \quad ,$$

protože momentová matice $X'X$ je sama (za přijatého předpokladu, že X má plnou hodnost k) vždy pozitivně definitní.

K získání Fisherovy informační matice (čtvercové symetrické matice řádu $k+1$) potřebujeme vypočítat derivace věrohodnostní funkce podle (složek) vektoru β a skaláru σ^2 . Vyjdeme-li ze vztahů

$$(A1) \quad \frac{\partial L^{**}}{\partial \beta} = \frac{1}{2\sigma^2} \cdot (-2X'y + 2X'X\beta) = 0$$

$$(B1) \quad \frac{\partial L^{**}}{\partial \sigma^2} = \frac{T}{2\sigma^2} - \frac{1}{2\sigma^4} (y - X\beta)'(y - X\beta) = 0$$

určíme potřebné druhé parciální derivace pro dosazení do matice

$$P = \begin{pmatrix} \frac{\partial^2 L^{**}(\beta, \sigma^2)}{\partial \beta \partial \beta} & \frac{\partial^2 L^{**}(\beta, \sigma^2)}{\partial \beta \partial \sigma^2} \\ \frac{\partial^2 L^{**}(\beta, \sigma^2)}{\partial \beta \partial \sigma^2} & \frac{\partial^2 L^{**}(\beta, \sigma^2)}{\partial \sigma^2 \partial \sigma^2} \end{pmatrix} = - \begin{pmatrix} \frac{\partial^2 L^*(\beta, \sigma^2)}{\partial \beta \partial \beta} & \frac{\partial^2 L^*(\beta, \sigma^2)}{\partial \beta \partial \sigma^2} \\ \frac{\partial^2 L^*(\beta, \sigma^2)}{\partial \beta \partial \sigma^2} & \frac{\partial^2 L^*(\beta, \sigma^2)}{\partial \sigma^2 \partial \sigma^2} \end{pmatrix}$$

¹ Tvaru, kdy po určení některého parametru (zde β) dosadíme získanou hodnotu do původní věrohodnostní funkce, abychom mohli určit další parametry (zde σ^2) se někdy říká *koncentrovaná věrohodnostní funkce*.

Výpočtem jednotlivých derivací dostáváme

$$(A2) \quad \frac{\partial^2 L^{**}}{\partial \beta \partial \beta} = \frac{X'X}{\sigma^2}$$

$$(B2) \quad \frac{\partial L^{**}}{\partial \sigma^2 \partial \sigma^2} = -\frac{T}{2\sigma^4} + \frac{2}{2\sigma^6}(y - X\beta)'(y - X\beta) = \frac{1}{\sigma^6} \varepsilon' \varepsilon - \frac{T}{2\sigma^4}$$

$$(AB2) \quad \frac{\partial L^{**}}{\partial \beta \partial \sigma^2} = \frac{-(X'X\beta - X'y)}{\sigma^4} = \frac{X'\varepsilon}{\sigma^4} \quad \text{a pro kontrolu}$$

$$(AB2) \quad \frac{\partial L^{**}}{\partial \sigma^2 \partial \beta} = -\frac{2(-X')(y - X\beta)}{2\sigma^4} = \frac{X'\varepsilon}{\sigma^4}$$

Matice P tedy nabude tvaru

$$P = \begin{pmatrix} \frac{X'X}{\sigma^2} & \frac{X'\varepsilon}{\sigma^4} \\ \frac{X'\varepsilon}{\sigma^4} & \frac{1}{\sigma^6} \varepsilon' \varepsilon - \frac{T}{2\sigma^4} \end{pmatrix} = \frac{1}{\sigma^2} \begin{pmatrix} X'X & \frac{X'\varepsilon}{\sigma^2} \\ \frac{X'\varepsilon}{\sigma^2} & \frac{1}{\sigma^2} \varepsilon' \varepsilon - \frac{T}{2\sigma^2} \end{pmatrix}$$

Uplatněním střední hodnoty na P dostaneme

$$EP = \begin{pmatrix} X'X & E \frac{X'\varepsilon}{\sigma^2} \\ E \frac{X'\varepsilon}{\sigma^2} & E \left(\frac{1}{\sigma^4} \varepsilon' \varepsilon - \frac{T}{2\sigma^2} \right) \end{pmatrix} = \frac{1}{\sigma^2} \begin{pmatrix} X'X & 0 \\ 0 & \frac{T}{2\sigma^2} \end{pmatrix}, \text{ protože}$$

$$E \frac{X'\varepsilon}{\sigma^2} = \frac{X'}{\sigma^2} E\varepsilon = 0 \quad \text{s ohledem na nestochastičnost } X' \text{ a centrování } \varepsilon \text{ a}$$

$$E \left(\frac{1}{\sigma^4} \varepsilon' \varepsilon - \frac{T}{2\sigma^2} \right) = \frac{E\varepsilon' \varepsilon}{\sigma^4} - \frac{T}{2\sigma^2} = \frac{T}{2\sigma^2}, \quad \text{protože } E\varepsilon' \varepsilon = T \cdot \sigma^2$$

Všimněme si, že tato matice je – s ohledem na pozitivní definitnost momentové matice $X'X$ - rovněž pozitivně definitní. Tím jsme dokázali, že jde skutečně o minimum (námi záporně vzaté) resp. o maximum (původní) věrohodnostní funkce.

Nyní se můžeme přesvědčit, zda jde skutečně o nejlepší odhady, což zjistíme vyčíslením Fisherovy informační matice, jež je inverzní k matici P.

$$\left(\frac{\partial^2 L^{**}(\theta)}{\partial \theta \partial \theta'} \right)^{-1} = \begin{pmatrix} \frac{X'X}{\sigma^2} & 0 \\ 0 & \frac{T}{2\sigma^4} \end{pmatrix}^{-1} = \sigma^2 \cdot \begin{pmatrix} X'X & 0 \\ 0 & \frac{T}{2\sigma^2} \end{pmatrix}^{-1}$$

θ je užito z důvodu úsporného značení, v našem případě $\theta = (\beta_1, \beta_2, \dots, \beta_k, \sigma^2)$.

Poznámka a současně definice:

Odhadová funkce $\hat{\theta}$ se nazývá **MVB (minimum variance bound) estimátor**, jestliže je nestranná a jestliže její kovarianční matice má velikost danou jako

$$\text{Cov}V = -[T.P(\theta)]^{-1}, \text{ kde } P(\theta) = E\left[\frac{\partial^2 \ln f(Z_t, \theta)}{\partial \theta \partial \theta'}\right],$$

kde $f(Z_t, \theta)$ je sdružená hustota (resp. věrohodnostní funkce $L(\theta, Z_t)$) rozdělení náhodného vektoru Z_t sdružujícího (v našem případě) pozorované hodnoty y_t, X_t

Odtud můžeme určit kovarianční matici odhadové funkce $\hat{\theta}$.

$$\text{Cov}V = -[T.P(\theta)]^{-1} = -T \left[E. \frac{\partial^2 L^*(\theta)}{\partial \theta \partial \theta'} \right]^{-1} = T \cdot \begin{pmatrix} X'X & 0 \\ \sigma^2 & T \\ 0 & \frac{T}{2\sigma^4} \end{pmatrix}^{-1} = \sigma^2 \begin{pmatrix} X'X & 0 \\ T & 1 \\ 0 & \frac{1}{2\sigma^2} \end{pmatrix}^{-1}$$

Odtud je vidět, že dolní hranice velikosti asymptotické kovarianční matice vektoru β libovolného estimátoru je dána výrazem $\text{Cov}(\hat{\beta}) = \sigma^2 \cdot \lim \left(\frac{X'X}{T} \right)^{-1}$

To však přesně odpovídá tvaru asymptotické kovarianční matice $\sqrt{T}(\hat{\beta} - \beta)$.

Současně je vidět, že dolní hranice pro rozptyl $\hat{\sigma}^2$ (získaná libovolným estimátorem) je dána jako

$$\text{var } \hat{\sigma}^2 = 2\sigma^4$$

Poznámka: Mezi odhady reziduálního rozptylu pořizovanými metodou nejmenších čtverců a metodou maximální věrohodnosti platí vztah :

$${}_{ML} \hat{\sigma}^2 = \frac{T-k}{T} {}_{OLS} \hat{\sigma}^2$$

Poznámka: Z předchozího je vidět, že odhad reziduálního rozptylu metodou maximální věrohodnosti není nestranný (zůstává však – stejně jako odhad prostou metodou nemenších čtverců OLS – konzistentní a asymptoticky nestranný)

Obecněji, ne však zcela univerzálně lze říci, že zatímco OLS-metody inklinují z hlediska svých vlastností k *nestranným odhadům* (které jsou *vydatné* jen za velmi vzácných okolností), pak ML-techniky poskytují *odhady vydatné* (obvykle na samé mezi možností daných *Cramér-Raovou dolní hranicí*), avšak *nestrannost* je vlastností, kterou od nich obvykle nelze očekávat. Ve víceroznicových regresních modelech ovšem ani simultánní OLS odhadové techniky neposkytují *nestranné odhady* (opět nepočítaje výjimky), takže tam dvěma aspoň požadovanými statistickými vlastnostmi zůstává jen *konzistence a asymptotická normalita*.

