Základy ekonometrie

III. Regresní model s jedinou vysvětlující proměnnou

Obsah tématu

Jednoduchý regresní model

- Bez úrovňové konstanty.
- Nepotřebujeme maticovou algebru.

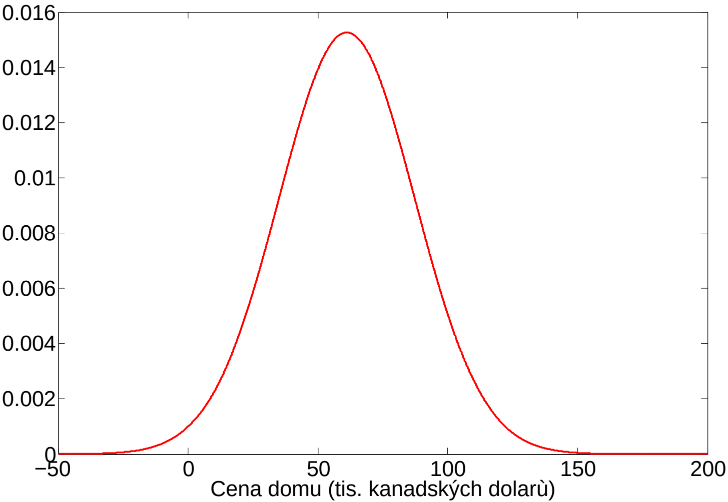
$$Y_i = \beta X_i + \epsilon_i.$$

Obsah tématu

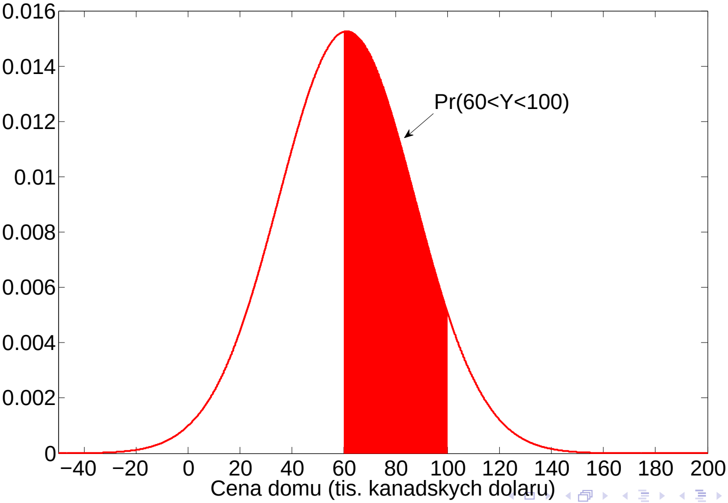
Náhodnost vysvětlované veličiny

- Vyjádření nejistoty – funkce hustoty pravděpodobnosti.
- Ilustrace na příkladu cen domů: dům o rozloze 5000 čtverečních stop.

Náhodnost vysvětlované veličiny – ilustrace funkce hustoty



Náhodnost vysvětlované veličiny – ilustrace oblasti pravděpodobnosti



Standardizace a Z-skór

- $Y \sim N(\mu, \sigma^2)$ a nová náhodná veličina:

$$Z = \frac{Y - \mu}{\sigma}.$$

- Střední hodnota a rozptyl Z:

$$\begin{aligned} E(Z) &= E\left(\frac{Y-\mu}{\sigma}\right) = \frac{E(Y-\mu)}{\sigma} \\ &= \frac{E(Y)-\mu}{\sigma} = \frac{\mu-\mu}{\sigma} = 0; \\ var(Z) &= var\left(\frac{Y-\mu}{\sigma}\right) = \frac{var(Y-\mu)}{\sigma^2} \\ &= \frac{var(Y)}{\sigma^2} = \frac{\sigma^2}{\sigma^2} = 1. \end{aligned}$$

- *Z-skór*: $Z \sim N(0, 1)$.

Standardizace a Z-skór – příklad

- Příklad z obrázku: $Y \sim N(61.153, 683.812)$.
- Barevná oblast: $\Pr(60 \leq Y \leq 100)$.

$$
\begin{aligned}
\Pr(60 \leq Y \leq 100) &= \Pr\left(\frac{60-\mu}{\sigma} \leq \frac{Y-\mu}{\sigma} \leq \frac{100-\mu}{\sigma}\right) \\
&= \Pr\left(\frac{60-61.153}{\sqrt{683.812}} \leq \frac{Y-61.153}{\sqrt{683.812}} \leq \frac{100-61.153}{\sqrt{683.812}}\right) \\
&= \Pr\left(-0.04 \leq Z \leq 1.49\right).
\end{aligned}
$$

- Podle toho, co nabízejí tabulky: např. hodnota distribuční funkce v 1.49 mínus hodnota v -0.04 (výsledek vždy stejný, 0.4479).

Obsah tématu

Klasické předpoklady v kontextu závisle proměnné

1. $E(Y_i) = \beta X_i$.
2. $var(Y_i) = \sigma^2$.
3. $cov(Y_i, Y_j) = 0$ pro $i \neq j$.
4. Y_i má normální rozdělení.
5. X_i je pevně dáno (nenáhodná veličina).

- Kompaktní vyjádření:

$$Y_i \sim N(\beta X_i, \sigma^2),$$

pro $i = 1, \ldots, N$, kdy Y_i a Y_j jsou vzájemně nekorelovány (pro $i \neq j$).

Očekávaná hodnota závisle proměné

- $E(Y_i) = \beta X_i$: předpoklad linearity modelu
- Model vícenásobné regrese: $E(Y_i) = \alpha + \beta_1 X_{1i} + \beta_2 X_{2i} + \ldots + \beta_k X_{ki}$.
- Očekáváme (v průměru), že Y_i leží na regresní přímce.

Konstantní rozptyl závisle proměné

- V realitě často porušen (příklad ceny malých a velkých domů nebo analýza spotřeby chudších a bohatších domácností).
- Splnění předpokladu – homoskedasticita.
- Porušení předpokladu – heteroskedasticita (použití zobecněné metody nejmenších čtverců).
- Úchvatný příspěvek v časopise *Econometrica*: angl. homoskedasticity $\times$ homoscedasticity.

Nekorelovanost pozorování závisle proměné

- Vzájemná nekorelovanost pozorování.

$$corr(Y_i, Y_j) = \frac{cov(Y_i, Y_j)}{\sqrt{var(Y_i)var(Y_j)}}$$

- $cov(Y_i, Y_j) = 0 \Leftrightarrow corr(Y_i, Y_j) = 0$.
- Obvykle splněno pro průřezová data – např. dotazníková šetření ($\times$ možnost prostorové korelace).
- V případě časových řad – obvykle problém (např. časová řada nezaměstnanosti).
- Korelace v čase – *autokorelace* (existence možných řešení).

Normalita rozdělení závisle proměné

- Motivace – obtížná.
- Předpoklad v empirických aplikacích obvyklý (ale ne vždy nutně reálný).
- Využití asymptotické teorie pro jeho uvolnění.

Nenáhodnost vysvětlující proměnné

- Experimentální vědy – dobrý předpoklad.
- V ekonomii – ne vždy vhodný předpoklad.
- Asymptotická teorie – vysvětlující proměnná náhodná $\times$ nekorelovaná s náhodnou složkou.
- V případě korelace vysvětlující veličiny a náhodné složky – OLS výsledky nekorektní (řešení v podobě metody instrumentálních proměnných).

Klasické předpoklady v kontextu náhodné složky

1. $E(\epsilon_i) = 0$.
2. $var(\epsilon_i) = E(\epsilon_i^2) = \sigma^2$. Konstantní rozptyl chyb (homoskedasticita).
3. $cov(\epsilon_i, \epsilon_j) = 0$ pro $i \neq j$. Náhodné složky jsou nekorelovány.
4. ϵ_i je normálně rozdělené.
5. X_i je pevně dáno (nenáhodná) veličina.

Nesplnění klasických předpokladů.

- Problém $\rightarrow$ zavádějící výsledky z OLS odhadu.
- Potřeba testování – obvykle na reziduích (odhady náhodných složek).
- Řešení – asymptotická teorie (velké vzorky) nebo jiné odhadové metody.

Obsah tématu

Estimátor a odhad

- Estimátor – obecný funkční předpis (funkce možných pozorování).
- Odhad – vyhodnocení funkce estimátoru na základě zjištěných pozorování.
- V klasické ekonometrii – neznámý parametr je pevně dané číslo $\times$ jeho odhad je realizace náhodné veličiny.

OLS estimátor

- Model bez úrovňové konstanty:

$$Y_i = \beta X_i + \epsilon_i .$$

- OLS estimátor na základě minimalizace součtu čtverců reziduí:

$$SSR = \sum_{i=1}^{N} \widehat{\epsilon}_i^2 \rightarrow \min .$$

- Výsledný vzorec – jednoduchý minimalizační problém $\Rightarrow$

$$\widehat{\beta} = \frac{\sum_{i=1}^{N} X_i Y_i}{\sum_{i=1}^{N} X_i^2} .$$

- $\widehat{\beta}$ – normálně rozdělená náhodná veličina (vztah závisí lineárně na Y_i = normální náhodné veličiny).

Alternativní vyjádření OLS estimátoru

- OLS estimátor – součet skutečné hodnoty parametru a části s chybovými členy a s hodnotami vysvětlujících proměnných.
- Výraz vhodný pro teoretická odvození (v praxi nevhodný – neznámé chybové členy).

$$\begin{aligned}\widehat{\beta} &= \frac{\sum X_i Y_i}{\sum X_i^2} = \frac{\sum X_i(X_i\beta + \epsilon_i)}{\sum X_i^2} \\ &= \beta + \frac{\sum X_i \epsilon_i}{\sum X_i^2}.\end{aligned}$$

Střední hodnota OLS estimátoru

- Nestranný (unbiased) estimátor – „v průměru“ roven skutečné hodnotě.

$$E\left(\widehat{\beta}\right) = \beta.$$

$$\begin{aligned}
E\left(\widehat{\beta}\right) &= E\left(\beta + \frac{\sum X_i \epsilon_i}{\sum X_i^2}\right) = \beta + E\left(\frac{\sum X_i \epsilon_i}{\sum X_i^2}\right) \\
&= \beta + \frac{1}{\sum X_i^2} E\left(\sum X_i \epsilon_i\right) = \beta + \frac{1}{\sum X_i^2} \sum X_i E(\epsilon_i) \\
&= \beta.
\end{aligned}$$

Rozptyl OLS estimátoru

- Význam variability (vydatnosti) estimátoru.

$$var\left(\widehat{\beta}\right) = \frac{\sigma^2}{\sum X_i^2}.$$

$$\begin{aligned} var\left(\widehat{\beta}\right) &= var\left(\beta + \frac{\sum X_i \epsilon_i}{\sum X_i^2}\right) = var\left(\frac{\sum X_i \epsilon_i}{\sum X_i^2}\right) \\ &= \left(\frac{1}{\sum X_i^2}\right)^2 var\left(\sum X_i \epsilon_i\right) = \left(\frac{1}{\sum X_i^2}\right)^2 \sum X_i^2 var(\epsilon_i) \\ &= \left(\frac{1}{\sum X_i^2}\right)^2 \sigma^2 \sum X_i^2 = \frac{\sigma^2}{\sum X_i^2}. \end{aligned}$$

Vydatnost estimátoru

- Celá řada nestranných estimátorů.
- Preferujeme ten nejvydatnější (nejlepší), tedy s nejmenším rozptylem $\rightarrow$ vydatný (efficient) estimátor.

Příklad méně vydatného nestranného estimátoru

- Pro jednoduchý regresní model bez úrovňové konstanty:

$$\tilde{\beta} = \frac{\sum_{i=1}^{N} Y_i}{\sum_{i=1}^{N} X_i}.$$

- Nestranný estimátor, ale vyšší rozptyl než OLS estimátor.
- Příklad – průměrná výška v populaci:
    - Lékař a náhodně vybraná skupina pacientů $\rightarrow$ výběrový průměr $\rightarrow$ nestranný odhad populační střední hodnoty (nejvydatnější).
    - Lékař a náhodně vybraná skupina pacientů $\rightarrow$ výška prvního z výběru $\rightarrow$ nestranný odhad populační střední hodnoty (ale vyšší rozptyl).

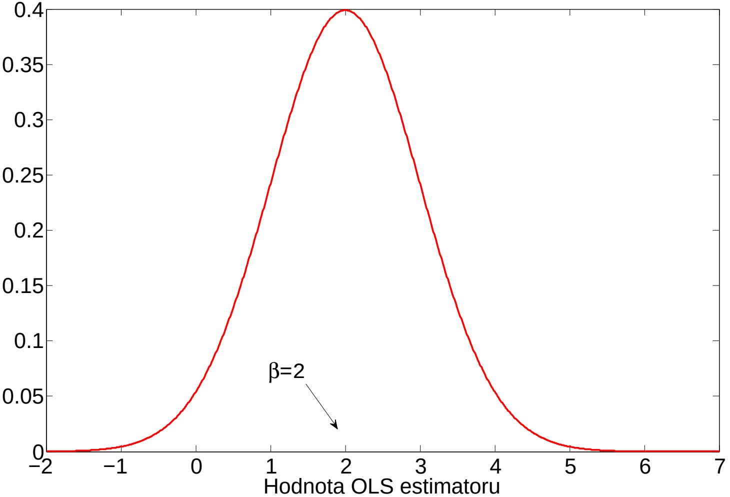
Rozdělení OLS estimátoru

- Při splnění klasických předpokladů:

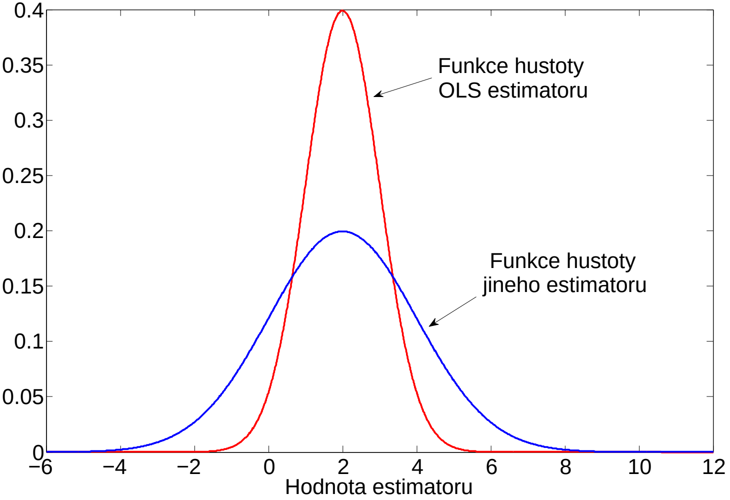
$$\widehat{\beta} \sim N\left(\beta, \frac{\sigma^2}{\sum X_i^2}\right).$$

- Důležité zejména pro odvození intervalů spolehlivosti a testování hypotéz.
- Příklad: $\widehat{\beta} \sim N(2,1)$ a $\tilde{\beta} \sim N(2,4)$.
- Ze statistických tabulek: např. $\Pr(1.0 \leq \widehat{\beta} \leq 3.0) = 0.68$ nebo $\Pr(0 \leq \widehat{\beta} \leq 1) = 0.14$.

Rozdělení OLS estimátoru (obrázek)



Porovnání OLS a jiného estimátoru (obrázek)



Gaussův-Markovův teorém

- Při splnění klasických předpokladů (normalita není nutná): OLS estimátor je BLUE (Best Linear Unbiased Estimator).
- Nejlepší, lineární, nestranný estimátor.

Carl Friedrich Gauss (1777–1855)

Andrej Markov (1856–1922)

Gaussův-Markovův teorém – nestrannost

- Jednoduchý regresní model bez úrovňové konstanty + splněny klasické předpoklady.
- Lineární estimátor: $\beta^* = c_1 Y_1 + \ldots + c_N Y_N$ pro konstanty $c_1, \ldots, c_N$.
- Nevychýlenost: $E(\beta^*) = \beta$.

$$
\begin{aligned}
E(\beta^*) &= E(c_1 Y_1 + \ldots + c_N Y_N) \\
&= c_1 E(Y_1) + \ldots + c_N E(Y_N) \\
&= c_1 \beta X_1 + \ldots + c_N \beta X_N \\
&= \beta \sum_{i=1}^{N} c_i X_i .
\end{aligned}
$$

- Nestranný estimátor, β^* – platí $\sum_{i=1}^{N} c_i X_i = 1$.

Gaussův-Markovův teorém – vydatnost

$$\begin{aligned} var\left(\beta^*\right) &= var\left(c_1 Y_1 + \ldots + c_N Y_N\right) \\ &= c_1^2 var\left(Y_1\right) + \ldots + c_N^2 var\left(Y_N\right) \\ &= c_1^2 \sigma^2 + \ldots + c_N^2 \sigma^2 \\ &= \sigma^2 \sum_{i=1}^{N} c_i^2. \end{aligned}$$

- Řešení problému: minimalizace rozptylu při omezení nestranností (předpoklad jednoduchého regresního modelu bez úrovňové konstanty a splnění klasických předpokladů).
- Hledáme $c_1, \ldots, c_N$ minimalizující $\sigma^2 \sum_{i=1}^{N} c_i^2$ při omezení $\sum_{i=1}^{N} c_i X_i = 1$.
- Standardní řešení (např. metoda Lagrangeových multiplikátorů).

Gaussův-Markovův teorém – řešení

- Řešení: $c_j = \dfrac{X_j}{\sum_{i=1}^{N} X_i^2}$, pro $j = 1, \ldots, N$.
- Dosazením do výrazu pro β^*:

$$\begin{aligned} \beta^* &= c_1 Y_1 + \ldots + c_N Y_N \\ &= \frac{X_1 Y_1}{\sum X_i^2} + \ldots + \frac{X_N Y_N}{\sum X_i^2} \\ &= \frac{\sum X_i Y_i}{\sum X_i^2} = \widehat{\beta}. \end{aligned}$$

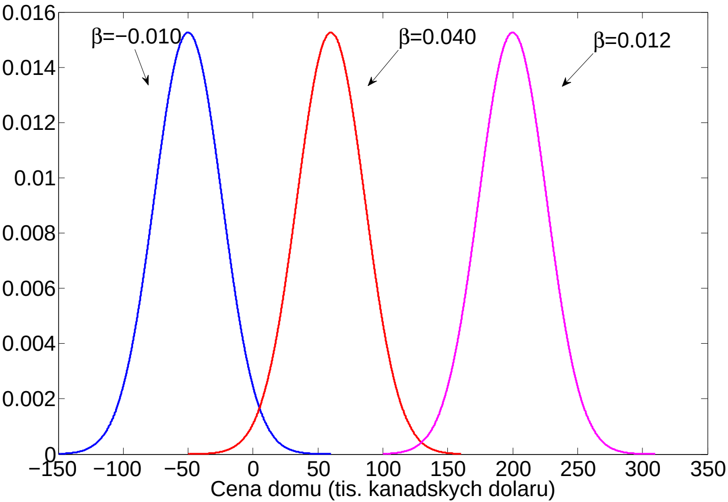
- Závěr: lineární, nestranný estimátor s minimálním rozptylem je estimátor metody nejmenších čtverců.

Obsah tématu

Motivace ML odhadu

- Maximal likelihood (ML) estimátor – velmi mocný ekonometrický nástroj.
- Funkce hustoty pravděpodobnosti pro Y by měla být v souladu s tím, co pozorujeme v datech.
- Příklad – ceny domů (pro různé hodnoty parametrů β).

Motivace ML odhadu – obrázek



Věrohodnostní funkce

- Věrohodnostní funkce: funkce sdružené hustoty pravděpodobnosti pro všechna pozorování.
- Při platnosti klasických předpokladů: $Y_1, \ldots, Y_N$ jsou normální, vzájemně nekorelované, náhodné veličiny (se střední hodnotou βX_i a rozptylem σ^2).
- Funkce hustoty pravděpodobnosti každé náhodné veličiny: $p(Y_i|X_i, \beta)$ (podmíněné hustoty).
- Sdružená hustota pravděpodobnosti všech pozorování:

$$p(Y_1, \ldots, Y_N) = \prod_{i=1}^{N} p(Y_i|X_i, \beta).$$

Princip ML odhadu

- Při známých hodnotách X_i a Y_i je věrohodnostní funkce funkcí parametrů – $L(\beta)$.
- Odhad metodou maximální věrohodnosti = nalezení takové hodnoty β, která maximalizuje věrohodnostní funkci $L(\beta)$ (zde analyticky, jinak numerická optimalizace).

$$p\left(Y_i|X_i\beta\right) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2\sigma^2}\left(Y_i - \beta X_i\right)^2\right].$$

$$\begin{aligned} L(\beta) &= \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{1}{2\sigma^2}\left(Y_i - \beta X_i\right)^2\right] \\ &= \frac{1}{\left(2\pi\sigma^2\right)^{\frac{N}{2}}} \exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{N}\left(Y_i - \beta X_i\right)^2\right]. \end{aligned}$$

Maximalizace věrohodnostní funkce.

- Obvykle práce s logaritmem věrohodnostní funkce, $l(\beta)$ (nemění se bod maxima).

$$\begin{aligned} l(\beta) &= \ln\left[L(\beta)\right] \\ &= \ln\left\{\frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}}\exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{N}(Y_i-\beta X_i)^2\right]\right\} \\ &= \ln\left\{\frac{1}{(2\pi\sigma^2)^{\frac{N}{2}}}\right\}-\frac{1}{2\sigma^2}\sum_{i=1}^{N}(Y_i-\beta X_i)^2. \end{aligned}$$

- Pro maximalizaci $l(\beta)$ potřeba minimalizovat člen $\sum_{i=1}^{N}(Y_i-\beta X_i)^2$ $\rightarrow$ estimátor metody nejmenších čtverců.

Obsah tématu

Interval spolehlivosti parametru

- $\widehat{\beta}$ má normální rozdělení se střední hodnotou β a danou hodnotou rozptylu.
- Konstrukce *Z-skóru*:

$$Z = \frac{\widehat{\beta} - E(\widehat{\beta})}{\sqrt{var(\widehat{\beta})}} = \frac{\widehat{\beta} - \beta}{\sqrt{\frac{\sigma^2}{\sum x_i^2}}}.$$

- $Z \sim N(0,1) \Rightarrow$ tabulky standardizovaného normálního rozdělení.
- Např.

$$\Pr\left[-1.96 \leq Z \leq 1.96\right] = 0.95.$$

Odvození intervalu spolehlivosti

- 95% interval spolehlivosti → stačí osamostatnit parametr β:

$$\Pr\left[-1.96 \le \frac{\widehat{\beta}-\beta}{\sqrt{\frac{\sigma^2}{\sum x_i^2}}} \le 1.96\right]$$
$$= \Pr\left[-1.96\sqrt{\frac{\sigma^2}{\sum x_i^2}} \le \widehat{\beta}-\beta \le 1.96\sqrt{\frac{\sigma^2}{\sum x_i^2}}\right]$$
$$= \Pr\left[\widehat{\beta}-1.96\sqrt{\frac{\sigma^2}{\sum x_i^2}} \le \beta \le \widehat{\beta}+1.96\sqrt{\frac{\sigma^2}{\sum x_i^2}}\right]$$
$$= 0.95.$$

Alternativní vyjádření

- 95% interval spolehlivosti:

$$\widehat{\beta} - 1.96\sqrt{\frac{\sigma^2}{\sum x_i^2}} \leq \beta \leq \widehat{\beta} + 1.96\sqrt{\frac{\sigma^2}{\sum x_i^2}}$$

- Alternativně:

$$\widehat{\beta} \pm 1.96\sqrt{\frac{\sigma^2}{\sum x_i^2}}$$

$$\left[\widehat{\beta} - 1.96\sqrt{\frac{\sigma^2}{\sum x_i^2}}, \widehat{\beta} + 1.96\sqrt{\frac{\sigma^2}{\sum x_i^2}}\right].$$

- Hodnota 95 %: *hladina spolehlivosti (confidence level)*.
- Analogicky při jiných hladinách spolehlivosti (např. 90 %) $\Rightarrow$ jiná čísla (kritické hodnoty) ze statistických tabulek (např. 1.64).
- Praktický problém: neznáme rozptyl chyb, σ^2.

Postup testování hypotéz

1. Specifikace nulové hypotézy, H_0 a alternativní hypotézy H_1.
2. Specifikace testové statistiky.
3. Nalezení rozdělení testové statistiky za předpokladu, že H_0 platí.
4. Volba hladiny významnosti.
5. Na základě kroků 3 a 4 získáme kritickou hodnotu.
6. Spočítáme testovou statistiku z kroku 2 a porovnáme ji s kritickou hodnotou z kroku 5:
   - Zamítáme H_0 (ve prospěch alternativní hypotézy), pokud absolutní hodnota testové statistiky je větší než kritická hodnota (v opačném případě H_0 nezamítáme) = spadá do *kritického oboru* ohraničeného kritickými hodnotami.

Testování hypotéz o parametru

1. $H_0 : \beta = \beta_0$, kdy β_0 je známo (obvykle $\beta_0 = 0$). Alternativní hypotéza je obvykle $H_1 : \beta \neq \beta_0$ (možná i jednostranná alternativa = jiná kritická hodnota).
2. $Z = \dfrac{\widehat{\beta} - \beta}{\sqrt{\dfrac{\sigma^2}{\sum X_i^2}}}$.
3. $Z = \dfrac{\widehat{\beta} - \beta_0}{\sqrt{\dfrac{\sigma^2}{\sum X_i^2}}} \sim N(0,1)$.
4. Obvyklá volba je 5 % (tj. 0.05).
5. Protože $Z \sim N(0,1)$ a $\Pr[-1.96 \leq Z \leq 1.96] = 0.95$, je kritická hodnota právě 1.96.
    - Oboustranný test hypotézy (stejně jako oboustranný interval spolehlivosti) $\rightarrow$ kritická hodnota = 97.5% (0.975) kvantilu standardizovaného normálního rozdělení $(1 - 0.05/2)$.
6. V našem případě zamítáme H_0, pokud $|Z| > 1.96$.

Estimátor pro rozptyl chyb

- Obvyklý estimátor pro σ^2 je výběrový rozptyl, s^2.
- OLS rezidua: $\widehat{\epsilon}_i = Y_i - \widehat{\beta}X_i$; OLS estimátor σ^2: $s^2 = \frac{\sum \widehat{\epsilon}_i^2}{N-1}$.
- Lze ukázat, že $E(s^2) = \sigma^2$.
- Neformální intuice: $E\left(\epsilon_i^2\right) = \sigma^2 \Rightarrow \epsilon_i^2$ by mohl být dobrý estimátor pro σ^2.
- Aritmetický průměr čtverců chyb jako estimátoru pro σ^2: $\dfrac{\sum \epsilon_i^2}{N}$.
- Nahrazení nepozorovatelných ϵ_i rezidui: $\frac{\sum \widehat{\epsilon}_i^2}{N}$ (vychýlený estimátor σ^2 jakožto výsledek metody maximální věrohodnosti).
- Nevychýlený OLS estimátor: N je nahrazeno $N-1$ (počet stupňů volnosti)
- Model vícenásobné regrese s k vysvětlujícími proměnnými a úrovňovou konstantu:

$$s^2 = \frac{\sum \widehat{\epsilon}_i^2}{N-k-1} = \frac{SSR}{N-k-1}.$$

Modifikace procedur při neznámén rozptylu chyb

- Konstrukce Z-skóru:

$$Z = \frac{\widehat{\beta} - \beta}{\sqrt{\frac{s^2}{\sum X_i^2}}} \sim t_{N-1},$$

- Testování hypotéz a testová statistika:

$$t = \frac{\widehat{\beta} - \beta_0}{\sqrt{\frac{s^2}{\sum X_i^2}}} \sim t_{N-1}.$$

- t_{N-1} je Studentovo t-rozdělení s $N-1$ stupni volnosti.
- Tabulky t-rozdělení nebo p-hodnota testu (zamítáme H_0 pokud p-hodnota je menší než hladina významnosti).

Obsah tématu

K čemu může být asymptotická teorie dobrá?

- Odvození vlastností estimátorů pro (nekonečně) velké vzorky.
- Odvození rozdělení testových statistik pro (nekonečně) velké vzorky u řady diagnostických testů.

Principy

- Co se děje, když velikost vzorku jde k nekonečnu?
- Žádný předpoklad na rozdělení náhodných složek.
- X_i je *nezávisle a identicky rozdělená* náhodná veličina *independent and identically distributed – i.i.d.*, nezávislá na náhodných složkách $\Rightarrow$ $E(X_i) = \mu_X$ a $var(X_i) = \sigma_X^2$.
- Používáme OLS estimátor v podobě:

$$\widehat{\beta} = \beta + \frac{\sum X_i \epsilon_i}{\sum X_i^2}.$$

Konzistence OLS estimátoru – úvod

- $\text{plim}(\widehat{\beta}) = \beta$.

$$
\begin{aligned}
\text{plim}\left(\widehat{\beta}\right) &= \text{plim}\left(\beta + \frac{\sum X_i \epsilon_i}{\sum X_i^2}\right) \\
&= \beta + \text{plim}\left(\frac{\sum X_i \epsilon_i}{\sum X_i^2}\right) \text{ dle Slutského teorému} \\
&= \beta + \text{plim}\left(\frac{\frac{1}{N}\sum X_i \epsilon_i}{\frac{1}{N}\sum X_i^2}\right) \\
&= \beta + \frac{\text{plim}\left(\frac{1}{N}\sum X_i \epsilon_i\right)}{\text{plim}\left(\frac{1}{N}\sum X_i^2\right)} \text{ dle Slutského teorému.}
\end{aligned}
$$

Konzistence OLS estimátoru – dokončení

- *Zákon velkých čísel* k vyhodnocení limity v pravděpodobnosti, $\text{plim}\left(\frac{1}{N}\sum X_i\epsilon_i\right)$.
- Podstata: *„průměr konverguje ke střední hodnotě."*
- Z předpokladu o nezávislosti chyb regrese a vysvětlující proměnné:

$$\text{plim}\left(\frac{1}{N}\sum X_i\epsilon_i\right) = E(X_i\epsilon_i) = 0$$

- Zákon velkých čísel $\rightarrow$ $\text{plim}\left(\frac{1}{N}\sum X_i^2\right) = E(X_i^2)$.
- Z definice rozptylu, $var(X_i) = E(X_i^2) - [E(X_i)]^2$, můžeme psát $E(X_i^2) = var(X_i) + [E(X_i)]^2 = \sigma_X^2 + \mu_X^2 \Rightarrow$

$$\text{plim}\left(\widehat{\beta}\right) = \beta + \frac{0}{\sigma_X^2 + \mu_X^2} = \beta.$$

Krátké zastavení na cestě k asymptotické normalitě

- Odvození jen na první pohled náročné a stresující.
- Pro testování hypotéz o parametru potřebujeme znát rozdělení OLS estimátoru pro nekonečně velké vzorky (asymptotické rozdělení).
- Následně využití i pro odvození pravděpodobnostního rozdělení jiných testových statistik u řady diagnostických testů (to již náročnější a stresující bývá $\Rightarrow$ nebudeme v tomto základním kurzu ekonometrie řešit).

Centrální limitní věta (základ odvození)

- Navazuje na zákon velkých čísel.
- Podstata centrální limitní věty (teorému): *„pravděpodobnostní rozdělení výběrového průměru konverguje v distribuci k normálně rozdělené náhodné veličině."*

Asymptotická normalita – úvod

- Pro $N \to \infty$ platí:

$$\sqrt{N}(\widehat{\beta} - \beta) \sim N\left(0, \frac{\sigma^2}{\sigma_X^2 + \mu_X^2}\right).$$

- Lze psát:

$$\sqrt{N}(\widehat{\beta} - \beta) = \sqrt{N}\frac{\sum X_i \epsilon_i}{\sum X_i^2} = \sqrt{N}\frac{\frac{1}{N}\sum X_i \epsilon_i}{\frac{1}{N}\sum X_i^2}.$$

- *Centrální limitní teorém* pro $N \to \infty$:

$$\sqrt{N}\frac{1}{N}\sum X_i \epsilon_i \sim N\left(0, var(X_i \epsilon_i)\right).$$

Asymptotická normalita – pokračování

$$
\begin{aligned}
var(X_i\epsilon_i) &= E(X_i^2\epsilon_i^2) - [E(X_i\epsilon_i)]^2 \\
&= E(X_i^2)E(\epsilon_i^2) - [E(X_i)E(\epsilon_i)]^2 \\
&= (\sigma_X^2 + \mu_X^2)\sigma^2 - [\mu_X 0]^2 \\
&= (\sigma_X^2 + \mu_X^2)\sigma^2.
\end{aligned}
$$

- Ukázali jsme si:

$$\text{plim}\left(\frac{1}{N}\sum X_i^2\right) = (\sigma_X^2 + \mu_X^2).$$

Asymptotická normalita – dokončení

- S využitím Cramerova teorému zkombinujeme výsledky centrální limitní věty s výrazy pro *var* $(X_i \epsilon_i)$ a plim $\left(\frac{1}{N}\sum X_i^2\right)$:

$$\sqrt{N}(\widehat{\beta} - \beta) \xrightarrow{N} N\left(0, \frac{(\sigma_X^2 + \mu_X^2)\sigma^2}{(\sigma_X^2 + \mu_X^2)^2}\right).$$

- Vzájemným vykrácením členů:

$$\sqrt{N}(\widehat{\beta} - \beta) \xrightarrow{N} N\left(0, \frac{\sigma^2}{(\sigma_X^2 + \mu_X^2)}\right).$$

Praktické využití

- Pro $N \to \infty$ konverguje $\sqrt{N}(\widehat{\beta} - \beta)$ k normálnímu rozdělení.
- Využitím vlastností operátorů střední hodnoty a rozptylu:

$$\widehat{\beta} \sim N\left(\beta, \frac{\sigma^2}{N(\sigma_X^2 + \mu_X^2)}\right).$$

- Problém: $(\sigma_X^2 + \mu_X^2)$ neznámé.
- Platí plim $\left(\frac{1}{N}\sum X_i^2\right) = \sigma_X^2 + \mu_X^2 \Rightarrow \left(\frac{1}{N}\sum X_i^2\right)$ je konzistentní estimátor pro $\sigma_X^2 + \mu_X^2$.
- Závěr:

$$\widehat{\beta} \sim N\left(\beta, \frac{\sigma^2}{\sum X_i^2}\right).$$

- Výsledky platí aproximativně!