Ekonometrija 1

Druge vaje:
Ponovitev osnov matrične algebre. Izpeljava cenilke regresijskih koeficientov po metodi najmanjših kvadratov in njene lastnosti.

Na drugih vajah bomo najprej ponovili osnove matrične algebre, nato pa še izpeljavo cenilke regresijskih koeficientov po metodi najmanjših kvadratov. Prikazali bomo preprost primer njene uporabe ter njene lastnosti.

♣-♣-♣

Primer 1: Izpeljite obrazec za izračun ocen regresijskih koeficientov po metodi najmanjših kvadratov in prikažite njegovo uporabo na primeru ocenjevanja regresijske funkcije:

$$y_i = \beta_1 + \beta_2 x_{2i} + \beta_3 x_{3i} + u_i,$$

pri čemer so za deset enot že izračunane naslednje količine:

$$\bar{y} = 2{,}4; \quad \bar{x}_2 = 0{,}4; \quad \bar{x}_3 = 0{,}8; \quad \sum y_i^2 = 184; \quad \sum x_{2i} x_{3i} = -2;$$
$$\sum x_{2i}^2 = 10; \quad \sum x_{3i}^2 = 12; \quad \sum y_i x_{2i} = 40; \quad \sum y_i x_{3i} = -2.$$

Izpeljava cenilke regresijskih koeficientov po metodi najmanjših kvadratov

Linearni populacijski regresijski model z dvema pojasnjevalnima spremenljivkama x_2 in x_3 zapišemo takole:

$$\begin{aligned} y_i &= \beta_1 + \beta_2 x_{2i} + \beta_3 x_{3i} + u_i, \\ E(y_i | x_{2i}, x_{3i}) &= \beta_1 + \beta_2 x_{2i} + \beta_3 x_{3i}. \end{aligned} \tag{1}$$

Ker vrednosti parametrov zgornjega populacijskega modela niso poznane, moramo določiti pravila njihovega ocenjevanja z uporabo vzorčnih podatkov. Zanje lahko, analogno z (1), oblikujemo naslednji linearni regresijski model:

$$\begin{aligned} y_i &= b_1 + b_2 x_{2i} + b_3 x_{3i} + e_i, \\ \hat{y}_i &= b_1 + b_2 x_{2i} + b_3 x_{3i}. \end{aligned} \tag{2}$$

Med različnimi možnimi kriteriji izberemo metodo (navadnih) najmanjših kvadratov (angl. *Ordinary Least Squares – OLS*), na podlagi katere ocenimo regresijske koeficiente populacijske regresijske funkcije tako, da je vsota kvadratov ostankov modela (2) najmanjša možna. Glede na to, da je posamezni ostanek opredeljen kot razlika med opazovano in ocenjeno vrednostjo odvisne spremenljivke, je vsota njihovih kvadratov linearna funkcija cenilk regresijskih koeficientov:

$$
\begin{gathered}
e_i = y_i - \hat{y}_i = y_i - b_1 - b_2 x_{2i} - b_3 x_{3i}, \\
\sum_{i=1}^{n} e_i^{\,2} = \sum_{i=1}^{n} (y_i - b_1 - b_2 x_{2i} - b_3 x_{3i})^2 = S(b_1, b_2, b_3).
\end{gathered}
\tag{3}
$$

Zapisano funkcijo najprej razčlenimo

$$
\begin{gathered}
S(b_1, b_2, b_3) = \sum_{i=1}^{n} (y_i - b_1 - b_2 x_{2i} - b_3 x_{3i})^2 = \sum_{i=1}^{n} y_i^2 + n b_1^2 + b_2^2 \sum_{i=1}^{n} x_{2i}^2 + \\
+ b_3^2 \sum_{i=1}^{n} x_{3i}^2 - 2b_1 \sum_{i=1}^{n} y_i - 2b_2 \sum_{i=1}^{n} x_{2i} y_i - 2b_3 \sum_{i=1}^{n} x_{3i} y_i + \\
+ 2b_1 b_2 \sum_{i=1}^{n} x_{2i} + 2b_1 b_3 \sum_{i=1}^{n} x_{3i} + 2b_2 b_3 \sum_{i=1}^{n} x_{2i} x_{3i},
\end{gathered}
\tag{4}
$$

nato pa parcialno odvajamo po b_j:

$$
\begin{gathered}
\frac{\partial S}{\partial b_1} = 2nb_1 + 2b_2 \sum_{i=1}^{n} x_{2i} + 2b_3 \sum_{i=1}^{n} x_{3i} - 2\sum_{i=1}^{n} y_i \ , \\
\frac{\partial S}{\partial b_2} = 2b_1 \sum_{i=1}^{n} x_{2i} + 2b_2 \sum_{i=1}^{n} x_{2i}^2 + 2b_3 \sum_{i=1}^{n} x_{2i} x_{3i} - 2\sum_{i=1}^{n} x_{2i} y_i \ , \\
\frac{\partial S}{\partial b_3} = 2b_1 \sum_{i=1}^{n} x_{3i} + 2b_2 \sum_{i=1}^{n} x_{2i} x_{3i} + 2b_3 \sum_{i=1}^{n} x_{3i}^2 - 2\sum_{i=1}^{n} x_{3i} y_i \ .
\end{gathered}
\tag{5}
$$

Če parcialne odvode izenačimo z nič in delimo z dva, dobimo naslednji sistem normalnih enačb:

$$
\begin{gathered}
b_1 n + b_2 \sum_{i=1}^{n} x_{2i} + b_3 \sum_{i=1}^{n} x_{3i} = \sum_{i=1}^{n} y_i, \\
b_1 \sum_{i=1}^{n} x_{2i} + b_2 \sum_{i=1}^{n} x_{2i}^2 + b_3 \sum_{i=1}^{n} x_{2i} x_{3i} = \sum_{i=1}^{n} x_{2i} y_i, \\
b_1 \sum_{i=1}^{n} x_{3i} + b_2 \sum_{i=1}^{n} x_{2i} x_{3i} + b_3 \sum_{i=1}^{n} x_{3i}^2 = \sum_{i=1}^{n} x_{3i} y_i,
\end{gathered}
\tag{6}
$$

ki ga lahko zapišemo tudi takole:

$$
\begin{bmatrix}
n & \sum_{i=1}^{n} x_{2i} & \sum_{i=1}^{n} x_{3i} \\
\sum_{i=1}^{n} x_{2i} & \sum_{i=1}^{n} x_{2i}^2 & \sum_{i=1}^{n} x_{2i} x_{3i} \\
\sum_{i=1}^{n} x_{3i} & \sum_{i=1}^{n} x_{2i} x_{3i} & \sum_{i=1}^{n} x_{3i}^2
\end{bmatrix}
\begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix}
=
\begin{bmatrix}
\sum_{i=1}^{n} y_i \\
\sum_{i=1}^{n} x_{2i} y_i \\
\sum_{i=1}^{n} x_{3i} y_i
\end{bmatrix}.
\tag{7}
$$

Če opredelimo:

$$\mathbf{X} = \begin{bmatrix} 1 & x_{21} & x_{31} \\ 1 & x_{22} & x_{32} \\ \vdots & \vdots & \vdots \\ 1 & x_{2n} & x_{3n} \end{bmatrix}, \quad \mathbf{X}^{\mathrm{T}} = \begin{bmatrix} 1 & 1 & \dots & 1 \\ x_{21} & x_{22} & \dots & x_{2n} \\ x_{31} & x_{32} & \dots & x_{3n} \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \\ b_3 \end{bmatrix}, \quad \mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \tag{8}$$

lahko ugotovimo, da velja:

$$\mathbf{X}^{\mathrm{T}}\mathbf{X} = \begin{bmatrix} n & \sum_{i=1}^{n} x_{2i} & \sum_{i=1}^{n} x_{3i} \\ \sum_{i=1}^{n} x_{2i} & \sum_{i=1}^{n} x_{2i}^2 & \sum_{i=1}^{n} x_{2i}x_{3i} \\ \sum_{i=1}^{n} x_{3i} & \sum_{i=1}^{n} x_{2i}x_{3i} & \sum_{i=1}^{n} x_{3i}^2 \end{bmatrix} \quad \text{in} \quad \mathbf{X}^{\mathrm{T}}\mathbf{y} = \begin{bmatrix} \sum_{i=1}^{n} y_i \\ \sum_{i=1}^{n} x_{2i}y_i \\ \sum_{i=1}^{n} x_{3i}y_i \end{bmatrix}. \tag{9}$$

To pomeni, da sistem normalnih enačb (6) oz. (7) lahko zapišemo v naslednji matrični obliki:

$$\mathbf{X}^{\mathrm{T}}\mathbf{X}\mathbf{b} = \mathbf{X}^{\mathrm{T}}\mathbf{y}. \tag{10}$$

Rešitev sistema linearnih enačb dobimo tako, da obe strani gornje matrične enačbe pomnožimo z inverzno matriko $(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}$:

$$(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}(\mathbf{X}^{\mathrm{T}}\mathbf{X})\mathbf{b} = (\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{y} \quad \Rightarrow \quad \mathbf{b} = (\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{y}. \tag{11}$$

■

Primer 2: Na podlagi vzorčnih podatkov za 10 enot smo ocenjevali linearno populacijsko regresijsko funkcijo $y_i = \beta_1 + \beta_2 x_{2i} + \beta_3 x_{3i} + u_i$ in dobili naslednje ocenjene vrednosti odvisne spremenljivke in ostanke regresijske funkcije vzorčnih podatkov:

i	1	2	3	4	5	6	7	8	9	10
$\hat{y}_i$	5,5	5	8	5,5	6,5	a	6	4	5,5	4,25
e_i	0,7	–2	0,5	1	–1	1	–1,5	0,5	b	1,3

Izračunajte manjkajoči vrednosti, ki sta označeni z a in b. Pri tem si pomagajte z značilnostmi cenilke regresijskih koeficientov po metodi najmanjših kvadratov.

Veljavnost osnovnih štirih značilnosti metode najmanjših kvadratov tudi dokažite.

■