Metoda Najmniejszych Kwadratów - rozwiązania.

Ekonometria WNE UW

Sebastian Zalas

s.zalas@uw.edu.pl

1. Pokaż że:

   (i) $\boldsymbol{X}'\boldsymbol{e} = 0$.

   Rozwiązanie: Rozpocznijmy od układu równań normalnych i skorzystajmy z faktu, że $\boldsymbol{y} = \boldsymbol{X}\hat{\boldsymbol{\beta}} + \boldsymbol{e}$.

$$
\begin{aligned}
(\boldsymbol{X}'\boldsymbol{X})\hat{\boldsymbol{\beta}} &= \boldsymbol{X}'\boldsymbol{y} \\
(\boldsymbol{X}'\boldsymbol{X})\hat{\boldsymbol{\beta}} &= \boldsymbol{X}'(\boldsymbol{X}\hat{\boldsymbol{\beta}} + \boldsymbol{e}) \\
(\boldsymbol{X}'\boldsymbol{X})\hat{\boldsymbol{\beta}} &= \boldsymbol{X}'\boldsymbol{X}\hat{\boldsymbol{\beta}} + \boldsymbol{X}'\boldsymbol{e} \\
\boldsymbol{X}'\boldsymbol{e} &= 0
\end{aligned}
$$

   Komentarz: $\boldsymbol{X}'\boldsymbol{e} = 0$ implikuje, że reszty $\boldsymbol{e}$ są nieskorelowane z regresorami $\boldsymbol{X}$. Zauważmy, że nie oznacza to, że składnik losowy $\boldsymbol{\epsilon}$ jest nieskorelowany z $\boldsymbol{X}$. Własność ta jest prawdziwa zarówno regresji ze stałą, jak i bez stałej.

   (ii) w prostym modelu ze stałą, suma reszt jest równa zero.

   Rozwiązanie:

$$
\begin{aligned}
\sum_i^n e_i &= 0 \\
\sum_i^n (y_i - \hat{y}_i) &= 0 \\
\sum_i^n (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) &= 0 \\
\sum_i^n (y_i - (\bar{y} - \hat{\beta}_1 \bar{x}) - \hat{\beta}_1 x_i) &= 0 \\
\sum_i^n (y_i - \bar{y} + \hat{\beta}_1 \bar{x} - \hat{\beta}_1 x_i) &= 0 \\
\sum_i^n y_i - \sum_i^n \bar{y} + \hat{\beta}_1 \sum_i^n \bar{x} - \hat{\beta}_1 \sum_i^n x_i &= 0 \\
n\bar{y} - n\bar{y} + \hat{\beta}_1 n\bar{x} - \hat{\beta}_1 n\bar{x} &= 0 \Rightarrow \sum_i^n e_i = 0
\end{aligned}
$$

   (iii) średnia wartość reszt jest równa zero.

   Rozwiązanie: Średnia reszt $\bar{e} = \frac{\sum_i e_i}{n}$. W poprzednim podpunkcie pokazaliśmy, że $\sum_i^n e_i = 0$, zatem $\bar{e} = 0$.

   (iv) prosta regresji przechodzi przez punkt $(\bar{x}, \bar{y})$.

   Rozwiązanie: Mamy pokazać, że prosta przechodzi przez punkt $(\bar{x}, \bar{y})$. Tak więc wystarczy podstawić punkt do równania prostej (oszacowanego modelu):

$$\begin{aligned}\hat{y} &= \hat{\beta}_0 + \hat{\beta}_1 x\\ \hat{y} &= \hat{\beta}_0 + \hat{\beta}_1 \bar{x}\\ \hat{y} &= \bar{y} - \hat{\beta}_1 \bar{x} + \hat{\beta}_1 \bar{x}\\ \hat{y} &= \bar{y}\end{aligned}$$

(v) wartości teoretyczne są nieskorelowane z resztami, czyli $\hat{\boldsymbol{y}}'\boldsymbol{e} = 0$.
Rozwiązanie: Skorzystajmy z faktu, że $\hat{\boldsymbol{y}} = \boldsymbol{X}\hat{\boldsymbol{\beta}}$.

$$\begin{aligned}\hat{\boldsymbol{y}}'\boldsymbol{e} &= 0\\ (\boldsymbol{X}\hat{\boldsymbol{\beta}})'\boldsymbol{e} &= 0\\ \hat{\boldsymbol{\beta}}'\boldsymbol{X}'\boldsymbol{e} &= 0\\ \hat{\boldsymbol{\beta}}'\boldsymbol{X}'\boldsymbol{e} &= 0\\ 0 &= 0\end{aligned}$$

Przypomnienie: $\boldsymbol{X}'\boldsymbol{e} = 0$.

2. Tabela zawiera wyniki ACT oraz GPA (grade point average) dla 8 studentów. GPA bazuje na 4-stopniowej skali i zostało zaokrąglone do jednego miejsca dziesiętnego.

Student	GPA	ACT
1	2.8	21
2	3.4	24
3	3.0	26
4	3.5	27
5	3.6	29
6	3.0	25
7	2.7	25
8	3.7	30

(i) Oszacuj zależność między GPA a ACT za pomocą MNK, tzn. poszukaj oszacowań wyrazu wolnego i nachylenia.
Rozwiązanie: Równanie modelu do oszacowania:

$$GPA = \beta_0 + \beta_1 ACT + \varepsilon$$

Musimy znaleźć $\hat{\beta}_0$ oraz $\hat{\beta}_1$ czyli oszacowania parametrów modelu. Skorzystajmy z wyprowadzonych wzorów:

$$\begin{aligned}\hat{\beta}_1 &= \frac{S_{GPA,ACT}}{S_{ACT}} = \frac{0.8303571}{8.125} = 0.1021978\\ \hat{\beta}_0 &= G\bar{P}A - \hat{\beta}_1 A\bar{C}T = 3.2125 - 0.1021978 \times 25.875 = 0.5681319\end{aligned}$$

Zapiszmy oszacowany model

$$GPA = 0.5681319 + 0.1021978 \times ACT + e$$

(ii) Skomentuj kierunek relacji. Czy stała ma sensowną interpretację? Wyjaśnij. O ile wzrośnie GPA jeśli wynik ACT wzrośnie o 5 punktów?

Rozwiązanie: Oszacowanie współczynnika β_1 jest dodatnie co oznacza, że im wyższy wynik testu ACT, tym wyższa średnia ocen.

Gdy wynik ACT wzrośnie o 5 punktów ($\Delta ACT = 5$), wtedy:

$$\begin{aligned}\Delta GPA &= \hat{\beta}_1 \Delta ACT \\ &= 0.1021978 \times 5 = 0.510989\end{aligned}$$

GPA wzrośnie o 0.510989. (Gdy liczymy przyrost, stała nie ma znaczenia).

(iii) Jaka będzie wartość teoretyczna GPA kiedy $ACT = 20$.

Rozwiązanie: $G\hat{P}A = 0.5681319 + 0.1021978 \times 20 = 2.612088$

(iv) Oblicz wartości teoretyczne oraz reszty dla każdej obserwacji, sprawdź czy suma reszt jest równa 0.

Rozwiązanie: Korzystamy z równianina regresji $e_i = GPA - G\hat{P}A = GPA - 0.5681319 - 0.1021978 \times ACT$:

Student	GPA	ACT	reszty
1	2.8	21	0.0857143
2	3.4	24	0.3791209
3	3.0	26	-0.2252747
4	3.5	27	0.1725275
5	3.6	29	0.0681319
6	3.0	25	-0.1230769
7	2.7	25	-0.4230769
8	3.7	30	0.0659341

Suma reszt jest równa w przybliżeniu zero, co zgadza się z własnością udowodnianą w zadaniu 1.

(v) Zapisz model w formie macierzowej, oszacuj parametry.

Rozwiązanie: Model w formie macierzowej:

$$\boldsymbol{GPA} = [\mathbf{1},\ \boldsymbol{ACT}] \times \boldsymbol{\beta} + \boldsymbol{e}$$

$$\begin{bmatrix} 2.8 \\ 3.4 \\ 3.0 \\ 3.5 \\ 3.6 \\ 3.0 \\ 2.7 \\ 3.7 \end{bmatrix} = \begin{bmatrix} 1 & 21 \\ 1 & 24 \\ 1 & 26 \\ 1 & 27 \\ 1 & 29 \\ 1 & 25 \\ 1 & 25 \\ 1 & 30 \end{bmatrix} \times \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \\ e_3 \\ e_4 \\ e_5 \\ e_6 \\ e_7 \\ e_8 \end{bmatrix}$$

Szacowanie modelu:

$$\hat{\boldsymbol{\beta}} = \left(\begin{bmatrix} 1 & 21 \\ 1 & 24 \\ 1 & 26 \\ 1 & 27 \\ 1 & 29 \\ 1 & 25 \\ 1 & 25 \\ 1 & 30 \end{bmatrix}' \begin{bmatrix} 1 & 21 \\ 1 & 24 \\ 1 & 26 \\ 1 & 27 \\ 1 & 29 \\ 1 & 25 \\ 1 & 25 \\ 1 & 30 \end{bmatrix} \right)^{-1} \begin{bmatrix} 1 & 21 \\ 1 & 24 \\ 1 & 26 \\ 1 & 27 \\ 1 & 29 \\ 1 & 25 \\ 1 & 25 \\ 1 & 30 \end{bmatrix}' \begin{bmatrix} 2.8 \\ 3.4 \\ 3.0 \\ 3.5 \\ 3.6 \\ 3.0 \\ 2.7 \\ 3.7 \end{bmatrix} = \begin{bmatrix} 0.5681319 \\ 0.1021978 \end{bmatrix}$$

Zauważ że uzyskane oszacowania są równe wcześniej otrzymanym oszacowaniom w podpunkcie (i).

Komentarz: W repozytorium znajduje się kod, który szacuje model w tm zadaniu, zarówno korzystając z obu sposobów.

3. Z pewnej regresji otrzymano następujący wynik:

$$\boldsymbol{X}'\boldsymbol{X} = \begin{bmatrix} 6 & 6 & 5 \\ 6 & 9 & 7 \\ 5 & 7 & 6 \end{bmatrix}, \ \boldsymbol{X}'\boldsymbol{y} = \begin{bmatrix} 2 \\ 4 \\ 2 \end{bmatrix}$$

Oblicz estymator $\hat{\beta}$ bezpośrednio z układu równań normalnych.

Rozwiązanie: Skorzystaj ze wzoru $\hat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}$. Możesz obrócić macierz w `R` korzystając z funkcji `solve()`.

$\hat{\boldsymbol{\beta}} = \begin{bmatrix} -3.330669e^{-16} \\ 2 \\ -2 \end{bmatrix}$

4. Badacz używając danych o rozmiarze klasy (CS) oraz średnim wyniku testu ze 100 klas, oszacował poniższe równanie:

$$Test\hat{S}core = 520.4 - 5.82 \times CS$$

$R^2 = 0.08$, $SER = 11.5$.

(i) Klasa liczy 22 uczniów. Jaka jest wartość przewidywana średniego wyniku testu?

Rozwiązanie: $Test\hat{S}core = 520.4 - 5.82 \times 22 = 392.36$

(ii) W ostatnim roku klasa liczyła 19 uczniów w tym roku liczy ona 23 studentów. Jaka jest wartość teoretyczna średniego wyniku testu?

Rozwiązanie: $\Delta Test\hat{S}core = -5.82 \times (23 - 19) = -23.28$

(iii) Średni rozmiar klasy w próbie 100 klas wynosi 21.4. Jaki jest średni wynik testu w próbie 100 klas?

Rozwiązanie: $Test\bar{S}core = 520.4 - 5.82 \times 21.4 = 395.852$

5. Wylosowano próbę 200-stu dwudziestoletnich mężczyzn z populacji. Zebrano dane o wzroście i wadze. Otrzymano poniższe oszacowanie:

$$w\hat{a}ga = -99.41 + 3.94 \times wzrost$$

$R^2 = 0.81$, $SER = 10.2$, gdzie $waga$ jest mierzona w funtach oraz $wzrost$ jest mierzony w calach.

(i) Jaka jest przewidywana waga według regresji dla kogoś kto ma 70 cali wzrostu? 65 cali? 74 cali?

Rozwiązanie: Np.: $\hat{waga} = -99.41 + 3.94 \times 70 = 176.39$

(ii) Jeden z mężczyzn rośnie 1.5 cala w ciągu roku. Jaki jest teoretyczny przyrost wagi?

Rozwiązanie: $\Delta \hat{waga} = 3.94 \times 1.5 = 5.91$

(iii) Zamiast mierzenia wagi w funtach i wzrostu w calach, zmienne zostały zmierzone w centymetrach oraz kilogramach. Jakie są oszacowania z nowej regresji?

Rozwiązanie na następnych zajęciach.