Ekonometria IiE

Jerzy Mycielski

16 kwietnia 2024

Normalizacja modeli dla zmiennych binarnych

- Model wyjściowy $Var(\varepsilon_i) = \sigma^2$:

$$y_i = \begin{cases} y_i = 0 \text{ dla } x_i\beta + \varepsilon_i < 0 \\ y_i = 1 \text{ dla } x_i\beta + \varepsilon_i \geq 0 \end{cases}$$

- Normalizacja modelu:

$$y_i = \begin{cases} y_i = 0 \text{ dla } x_i\frac{\beta}{\sigma} + \frac{\varepsilon_i}{\sigma} < 0 \\ y_i = 1 \text{ dla } x_i\frac{\beta}{\sigma} + \frac{\varepsilon_i}{\sigma} \geq 0 \end{cases}$$

- Te dwa modele są równoważne obserwacyjnie! Parametr σ jest nieidentyfikowalny.
- Oszacować można tylko model dla znormalizowanego $\sigma = 1$:

$$y_i = \begin{cases} y_i = 0 \text{ dla } x_i\beta^* + \varepsilon_i^* < 0 \\ y_i = 1 \text{ dla } x_i\beta^* + \varepsilon_i^* \geq 0 \end{cases}$$

Tablica klasyfikacyjna

- Liczby obserwacji, dla których przewidywany sukces ($\widehat{p}_i \geq p^*$) bądź porażka ($\widehat{p}_i < p^*$)
- p^* prawdopopodobieństwo progowe

	Zaobserwowano		
Prognozowane	0	1	Razem
0	n_{00}	n_{01}	$n_{00} + n_{01}$
1	n_{10}	n_{11}	$n_{10} + n_{01}$
Razem	$n_{00} + n_{10}$	$n_{01} + n_{11}$	n

Pseudo R^2 liczebnościowe

- Pseudo R^2 liczebnościowe:

$$R^2_{liczebno\acute{s}ciowe} = \frac{n_{00} + n_{11}}{n},$$

- Dla modelu jedynie ze stałą zarówno probit jak i logit osiągają najwyższą wartość funkcji wiarygodności, gdy prawdopodbieństwo liczniejszej kategorii było równe 1.
- Zatem dla modelu tylko ze stałą $R^2_{liczebno\acute{s}ciowe} = \frac{n_{\max}}{n} > 0.5$, gdzie $n_{\max}$ jest liczebnością liczniejszej kategorii.
- Skorygowane $R^2_{liczebno\acute{s}ciowe}$ koryguje ten problem.

$$\overline{R}^2_{liczebno\acute{s}ciowe} = \frac{n_{00} + n_{11} - n_{\max}}{n - n_{\max}}.$$

Wrażliwość, Specyficzność, Wartość predykcyjna

- Wrażliwość (czułość) to proporcja sukcesów prawidłowo przewidzianych do ogółu obserwacji

$$\text{czułość} = Pr\left(\widehat{p}_i \geq p^* | \, y_i = 1\right) \approx \frac{n_{11}}{n_{01} + n_{11}}$$

- Specyficzność (swoistość)

$$\text{swoistość} = Pr\left(\widehat{p}_i \leq p^* | \, y_i = 0\right) \approx \frac{n_{00}}{n_{00} + n_{10}}$$

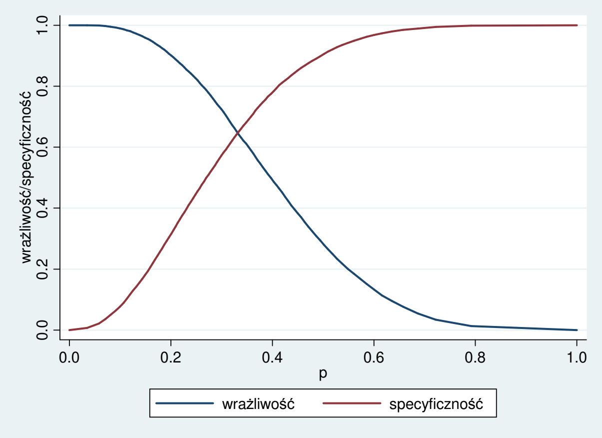
- Wartość predykcyjna

$$\text{Wartość predykcyjna dodatnia} = \frac{n_{11}}{n_{10} + n_{11}}$$

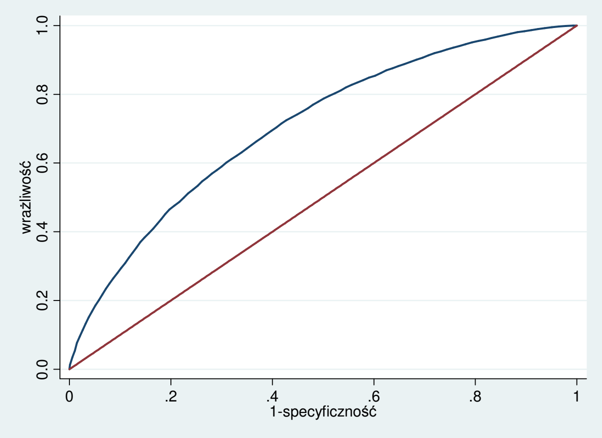
- Wartość predykcyjna ujemna

$$\text{Wartość predykcyjna ujemna} = \frac{n_{00}}{n_{00} + n_{01}}$$

Wrażliwość specyficzność



Krzywa ROC



- Standardowy wzór dla R^2 dla MNK można zapisać następująco

$$R^2 = \frac{ESS}{TSS} = \frac{ESS}{ESS + RSS} = \frac{\frac{1}{N}\sum_{i=1}^{N}\left(\widehat{y}_i - \overline{\widehat{y}}\right)^2}{\frac{1}{N}\sum_{i=1}^{N}\left(\widehat{y}_i - \overline{\widehat{y}}\right)^2 + \widetilde{\sigma}^2},$$

- Analogicznie dla zmiennej ukrytej: możemy zdefiniować:

$$R^2_{McKelvey,\ Zavoina} = \frac{\frac{1}{N}\sum_{i=1}^{N}\left(\widehat{y}_i^* - \overline{\widehat{y}}^*\right)^2}{\frac{1}{N}\sum_{i=1}^{N}\left(\widehat{y}_i^* - \overline{\widehat{y}}^*\right)^2 + \widetilde{\sigma}^2},$$

gdzie $\widehat{y}_i^* = x_i\widetilde{\beta}$.

Pseudo R^2 McKalveya i Zavoina

- Probit

$$R^2_{McKelvey,\ Zavoina} = \frac{\frac{1}{N}\sum_{i=1}^{N}\left(\widehat{y}_i^* - \overline{\widehat{y}}^*\right)^2}{\frac{1}{N}\sum_{i=1}^{N}\left(\widehat{y}_i^* - \overline{\widehat{y}}^*\right)^2 + 1}$$

- Logit

$$R^2_{McKelvey,\ Zavoina} = \frac{\frac{1}{N}\sum_{i=1}^{N}\left(\widehat{y}_i^* - \overline{\widehat{y}}^*\right)^2}{\frac{1}{N}\sum_{i=1}^{n}\left(\widehat{y}_i^* - \overline{\widehat{y}}^*\right)^2 + \frac{\pi^2}{3}}$$