Andrzej Kokoszkiewicz
Wyższa Szkoła Społeczno-Ekonomiczna w Warszawie

Współczynnik poprawności korelacyjnej

Niech zbiór $Z(m)$ zmiennych wyróżnionych w liniowym modelu ekonometrycznym zawiera zmienną objaśnianą Y oraz zmienne objaśniające $X_1, \ldots, X_m$. Niech $Z(m)$ generuje parę korelacyjną $(R(m), R_0(m))$, gdzie $R(m) = [r_{ij}]_{m \times m}$ jest macierzą korelacji, a $R_0(m) = [r_i]_{m \times 1}$ — wektorem korelacji, przy tym r_{ij} jest współczynnikiem korelacji między zmiennymi X_i, X_j, zaś r_i — współczynnikiem korelacji między zmiennymi Y, X_i, $i, j = 1, \ldots, m$.

Definiuje się

1) potencjał korelacyjny
$$C(m) = \left(\sum_{i,j=1}^{m} |r_{ij}|^2 \right)^{1/2}$$
macierzy $R(m)$ jako miarę wzajemnego skorelowania zmiennych objaśniających,

2) potencjał korelacyjny
$$C_0(m) = \left(\sum_{i=1}^{m} |r_i|^2 \right)^{1/2}$$
wektora $R_0(m)$ jako miarę skorelowania zmiennych objaśniających ze zmienną objaśnianą,

3) współczynnik poprawności korelacyjnej
$$K(m) = \frac{C_0(m)}{C(m)}$$
zbioru $Z(m)$ jako miarę relacji ilościowej pomiędzy skorelowaniem zmiennych objaśniających ze zmienną objaśnianą oraz skorelowaniem zmiennych objaśniających między sobą.

Im większa wartość $K(m)$, tym słabsze wzajemne skorelowanie zmiennych objaśniających i ich silniejsze skorelowanie ze zmienną objaśnianą, a więc wyższa jakość modelu.

Ponieważ $C(m) \in \left\langle \sqrt{m}, m \right\rangle$, $C_0(m) \in \left\langle 0, \sqrt{m} \right\rangle$, to $C_0(m) \leq C(m)$, $K(m) \in \langle 0, 1 \rangle$. $K(m)$ określa, jaka część skorelowania zmiennych objaśniających między sobą stanowi skorelowanie zmiennej objaśnianej ze zmiennymi objaśniającymi.

Stosując kryterium największego współczynnika $K(m)$ doboru zmiennych objaśniających do modelu otrzymuje się optymalny podzbiór tych zmiennych bez współliniowości.