MODEL REGRESJI WIELU ZMIENNYCH

Badając więcej aniżeli dwie cechy mierzalne w populacji posługujemy się pojęciem regresji wielorakiej, wielokrotnej, wielowymiarowej,

wg. Z. Pawłowskiego funkcją regresji 1-szego rodzaju zmiennej losowej Y względem zmiennych losowych X_1, X_2, ... , X_k jest wartość oczekiwana warunkowego rozkładu tej zmiennej $E(Y/X_1=x_1, \ldots, X_k=x_k)$

- Model regresji liniowej I rodzaju:

$$Y = \alpha_0 + \alpha_1 X_1 + \alpha_2 X_2 + \ldots + \alpha_k X_k + \xi$$

- Model regresji liniowej II rodzaju na podstawie danych empirycznych z próby losowej:

$$y = a_0 + \alpha_1 x_1 + a_2 x_2 + \ldots + a_k x_k + e_i$$

Funkcja regresji: $$\hat{y}_i = a_0 + \sum_{j=1}^{k} a_j x_j$$

Oszacowanie parametrów a_i powyższej funkcji regresji wiąże się z minimalizacją formy kwadratowej: $\sum (y_i - \hat{y}_i)^2 \to \min$

- W przypadku dwóch zmiennych egzogenicznych otrzymujemy następujący układ równań normalnych:

$$\begin{cases} \sum y = a_0 n + a_1 \sum x_1 + a_2 \sum x_2 \\ \sum y x_1 = a_0 \sum x_1 + a_1 \sum x_1^2 + a_2 \sum x_1 x_2 \\ \sum y x_2 = a_0 \sum x_2 + a_1 \sum x_1 x_2 + a_2 \sum x_2^2 \end{cases}$$

Układ powyższy można rozwiązać różnymi metodami, np. wyznacznikową, przez podstawienie itd.
Posługując się metodą wyznacznikową powyższy układ równań normalnych ma następujące rozwiązanie zgodne z KMNK.
Przyjmując oznaczenia:

$$Q = \begin{vmatrix} n & \sum x_1 & \sum x_2 \\ \sum x_1 & \sum x_1^2 & \sum x_1 x_2 \\ \sum x_2 & \sum x_1 x_2 & \sum x_2^2 \end{vmatrix}$$

otrzymujemy:

$$a_0 = \frac{\begin{vmatrix} \sum y & \sum x_1 & \sum x_2 \\ \sum y x_1 & \sum x_1^2 & \sum x_1 x_2 \\ \sum y x_2 & \sum x_1 x_2 & \sum x_2^2 \end{vmatrix}}{Q} = \frac{Q_0}{Q}$$

$$a_1 = \frac{\begin{vmatrix} n & \sum y & \sum x_2 \\ \sum x_1 & \sum y x_1 & \sum x_1 x_2 \\ \sum x_2 & \sum y x_2 & \sum x_2^2 \end{vmatrix}}{Q} = \frac{Q_1}{Q}$$

$$a_2 = \frac{\begin{vmatrix} n & \sum x_1 & \sum y \\ \sum x_1 & \sum x_1^2 & \sum y x_1 \\ \sum x_2 & \sum x_1 x_2 & \sum y x_2 \end{vmatrix}}{Q} = \frac{Q_2}{Q}$$

Parametry a_j funkcji regresji trzech zmiennych można też wyznaczyć w inny sposób (zgodny z KMNK), wg. tzw. wzorów Ezekiela:

$$a_1 = \frac{s(y)}{s(x_1)} \cdot \frac{r_{yx_1} - r_{yx_2} r_{x_1 x_2}}{1 - r_{x_1 x_2}^2}$$

$$a_2 = \frac{s(y)}{s(x_2)} \cdot \frac{r_{yx_2} - r_{yx_1} r_{x_1 x_2}}{1 - r_{x_1 x_2}^2}$$

$$a_0 = \frac{\sum y}{n} - \frac{a_1 \sum x_1}{n} - \frac{a_2 \sum x_2}{n} = \bar{y} - a_1 \bar{x}_1 - a_2 \bar{x}_2$$

- Parametry a_i nazywamy cząstkowymi współczynnikami regresji. Określają one średnią zmianę zmiennej Y na jednostkę przyrostu zmiennej niezależnej X_i przy założeniu stałego poziomu pozostałych zmiennych objasniających

- współczynniki β służące do określenia zmiany zmiennej zależnej wynikającej z jednostkowego wzrostu zmiennej niezależnej z uwzględnieniem relatywnej zmienności obu cech

$$\beta_1 = a_1 \frac{s(x_1)}{s(y)} \qquad \beta_2 = a_2 \frac{s(x_2)}{s(y)}$$

- w zapisie macierzowym wektor α szacujemy za pomocą estymatora a: $a = (X^T X)^{-1} X^T Y$

JAKOŚĆ DOPASOWANIA FUNKCJI REGRESJI WIELORAKIEJ określa się przy pomocy identycznych miar jak w przypadku dwóch zmiennych

- odchylenie standardowe składnika resztowego

$$S_e^2 = \frac{\sum (y_i - \hat{y}_i)^2}{n - m}$$

m – liczba szacowanych parametrów (liczba cząstkowych współczynników regresji i wyraz wolny) $m=k+1$

- współczynnik zmienności resztowej

$$V_{S_e} = \frac{S_e}{\bar{y}}$$

- współczynnik determinacji

$$R^2 = \left(1 - \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y})^2}\right)$$

$$R^2 = 1 - \frac{(n-k)S_e^2}{n \cdot S_y^2}$$

$$R^2 = 1 - \varphi^2$$

- współczynnik zgodności

$$\varphi^2 = \frac{\sum (y_i - \hat{y}_i)^2}{\sum (y_i - \bar{y})^2}$$

$$\varphi^2 = \frac{(n-k)S_e^2}{n \cdot S_y^2}$$