STATYSTYKA OPISOWA

Dr Alina Gleska

Instytut Matematyki WE PP

28 września 2018

Model potęgowy

$$y = \alpha x^{\beta} e$$

można sprowadzić poprzez zlogarytmowanie obu stron równania do postaci liniowej

$$ln(y) = ln(\alpha) + \beta ln(x) + lne,$$

co możemy zapisać jako

$$y^* = \alpha^* + \beta x^* + e^*.$$

Otrzymaliśmy model regresji liniowej logarytmów z wartości zmiennej Y względem logarytmów z wartości zmiennej X.

Po oszacowaniu MNK parametrów α^* i β oceny parametrów wyjściowego modelu są takie:

$$\alpha = \exp(\alpha^*), \quad \beta = \beta^*.$$

Parametr α określa teoretyczny poziom zmiennej zależnej Y przy jednostkowej wartości zmiennej niezależnej X, natomiast parametr β określa procentową zmianę wartości zmiennej zależnej spowodowaną zmianą wartości zmiennej niezależnej o 1% (właściwość tę nazywa się elastycznością zmiennej zależnej względem zmiennej niezależnej).

Model wykładniczy

$$y = \alpha\beta^x e$$

można sprowadzić poprzez zlogarytmowanie obu stron równania do postaci liniowej

$$ln(y) = ln(\alpha) + xln(\beta) + lne,$$

co możemy zapisać jako

$$y^* = \alpha^* + \beta^* x + e^*.$$

Otrzymaliśmy model regresji liniowej logarytmów z wartości zmiennej Y względem zmiennej X.

Po oszacowaniu MNK parametrów α^* i β^* oceny parametrów wyjściowego modelu są takie:

$$\alpha = \exp(\alpha^*), \quad \beta = \exp(\beta^*).$$

Parametr α określa teoretyczny poziom zmiennej zależnej Y w okresie bezpośrednio poprzedzającym pierwszy badany okres (czyli dla $X = 0$), natomiast β może posłużyć do określenia średniookresowego tempa zmian zmiennej Y: $r_g = (\beta - 1) \cdot 100\%$. Wielkość r_g jest stałą stopą zmian i określa względny przyrost zmiennej Y w przeliczeniu na stałą jednostkę czasową.

Wskaźniki korelacyjne Pearsona wykorzystuje się do mierzenia siły korelacji cech mierzalnych, niezależnie od kształtu zależności między nimi.
Podstawę do obliczenia wskaźników korelacyjnych stanowi:

- pogrupowanie wartości cech X i Y w formie tablicy dwudzielnej o wymiarach $r \times k$ (r - liczba wierszy, k - liczba kolumn),
- obliczenie wariancji ogólnej cechy Y lub/i wariancji ogólnej cechy X według wzorów:

$$S^2(X) = \frac{1}{n}\sum_{i=1}^{r}(x_i^0)^2 n_{i\bullet} - \overline{x}^2, \quad S^2(Y) = \frac{1}{n}\sum_{j=1}^{k}(y_j^0)^2 n_{\bullet j} - \overline{y}^2,$$

- obliczenie wariancji międzygrupowej cechy Y lub/i cechy X według wzorów:

$$S_m^2(X) = \frac{1}{n}\sum_{j=1}^{k}(\overline{x_j} - \overline{x})^2 n_{\bullet j}, \quad S_m^2(Y) = \frac{1}{n}\sum_{i=1}^{r}(\overline{y_i} - \overline{y})^2 n_{i\bullet},$$

gdzie $\overline{x}, \overline{y}$ to zwykłe średnie arytmetyczne obliczane według wzorów:

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{r} x_i^0 n_{i\bullet}, \quad \overline{y} = \frac{1}{n}\sum_{j=1}^{k} y_j^0 n_{\bullet j},$$

zaś $\overline{x}_j$ to j-ta średnia dla cechy X przy wartości y_j cechy Y:

$$\overline{x}_j = \frac{1}{n_{\bullet j}} \sum_{i=1}^{r} x_i^0 n_{ij} \quad \text{dla} \quad j = 1, 2, \ldots, k,$$

a $\overline{y}_i$ to i-ta średnia dla cechy Y przy wartości x_i cechy X:

$$\overline{y}_i = \frac{1}{n_{i\bullet}} \sum_{j=1}^{k} y_j^0 n_{ij} \quad \text{dla} \quad i = 1, 2, \ldots, r.$$

Wskaźniki korelacyjne Pearsona są dane wzorami:

- dla cechy X względem cechy Y:

$$r_{P(X,Y)} = \sqrt{\frac{S_m^2(X)}{S^2(X)}},$$

- dla cechy Y względem cechy X:

$$r_{P(Y,X)} = \sqrt{\frac{S_m^2(Y)}{S^2(Y)}}.$$

Wskaźniki korelacyjne Pearsona $r_{P(X,Y)}$ i $r_{P(Y,X)}$ przyjmują wartości z przedziału $[0,1]$ i nie wskazują kierunku korelacji. Wartość 0 przyjmują w przypadku, gdy brak jest związku między cechami, natomiast wartość 1, gdy między cechami istnieje zależność funkcyjna.
Na ogół są niesymetryczne, czyli $r_{P(X,Y)} \neq r_{P(Y,X)}$. Ponadto między wskaźnikami korelacyjnymi Pearsona a współczynnikiem korelacji liniowej Pearsona zachodzą nierówności:

$$r^2_{P(X,Y)} > r^2 \text{ i } r^2_{P(Y,X)} > r^2.$$

Mierniki krzywoliniowości służą do badania stopnia krzywoliniowości związku:

- cechy X względem cechy Y: $M_{K(X,Y)} = r^2_{P(X,Y)} - r^2$,
- cechy Y względem cechy X: $M_{K(Y,X)} = r^2_{P(Y,X)} - r^2$.

Mierniki krzywoliniowości również przyjmują wartości z przedziału $[0,1]$, przy czym jeśli wartość miernika krzywoliniowości jest mniejsza lub równa $0,2$, to związek jest prostoliniowy, a jeśli przekracza $0,2$, to pomiędzy cechami zachodzi związek krzywoliniowy.
PRZYKŁAD.

Współczynniki korelacji cząstkowej i wielorakiej służą do mierzenia siły związku między wieloma cechami X_j $(j = 1, 2, \ldots, m)$. Można tu wyróżnić dwie typowe sytuacje badawcze, polegające na określeniu:

- siły związku korelacyjnego jedynie między dwiema dowolnymi cechami X_{j_1} i X_{j_2} $(1 \leq j_1 \neq j_2 \leq m)$, z wyłączeniem wpływu na ten związek innych cech (korelacja cząstkowa),
- siły związku korelacyjnego między wybraną cechą X_{j_1} a zespołem innych cech (korelacja wieloraka).

Współczynniki korelacji cząstkowej obliczamy na podstawie wzoru

$$r_{jl\bullet\Omega} = \frac{-\tilde{R}_{jl}}{\sqrt{\tilde{R}_{jj}\tilde{R}_{ll}}},$$

gdzie pierwsze dwie litery jl w indeksie $jl \bullet \Omega$ oznaczają, że obliczamy współczynnik korelacji cząstkowej dla cech X_j i X_l z wyłączeniem działania cech o numerach ze zbioru $\Omega = \{1, 2, \ldots, j-1, j+1, \ldots, l-1, l+1, \ldots, m\}$, $\tilde{R}_{jl}$, $\tilde{R}_{jj}$, $\tilde{R}_{ll}$ są dopełnieniami algebraicznymi macierzy korelacji R o wymiarach $m \times m$, którą uzyskuje się obliczając współczynniki korelacji liniowej Pearsona dla każdej pary cech osobno.

Macierz R współczynników korelacji liniowej jest macierzą symetryczną, a jej elementami na głównej przekątnej są jedynki ($r_{jj} = 1$). Tak więc:

$$R = \begin{bmatrix} 1 & r_{12} & r_{13} & \dots & r_{1m} \\ r_{21} & 1 & r_{23} & \dots & r_{2m} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ r_{m1} & r_{m2} & r_{m3} & \dots & 1 \end{bmatrix}$$

Minorem $det(R_{jl})$ macierzy korelacji R nazywamy wyznacznik podmacierzy powstałej przez wykreślenie j-tego wiersza oraz l-tej kolumny z macierzy R, natomiast dopełnienie algebraiczne $\tilde{R}_{jl}$ macierzy korelacji R określamy wzorem:

Współczynniki korelacji cząstkowej

$$\begin{aligned}\tilde{R}_{jl} &= (-1)^{j+l}\det(R_{jl}) = \\ &= (-1)^{j+l}\det\begin{bmatrix} 1 & \dots & r_{1,l-1} & r_{1,l+1} & \dots & r_{1m} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ r_{j-1,1} & \dots & r_{j-1,l-1} & r_{j-1,l+1} & \dots & r_{j-1,m} \\ r_{j+1,1} & \dots & r_{j+1,l-1} & r_{j+1,l+1} & \dots & r_{j+1,m} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ r_{m1} & \dots & r_{m,l-1} & r_{m,l+1} & \dots & 1 \end{bmatrix}\end{aligned}$$

Dopełnienie algebraiczne $\tilde{R}_{jj}$ jest określone poprzez minor macierzy korelacji R powstały przez wykreślenie j-tego wiersza i j-tej kolumny z macierzy R, a dopełnienie algebraiczne $\tilde{R}_{ll}$ – poprzez wykreślenie l-tego wiersza i l-tej kolumny z macierzy R.

Współczynniki korelacji cząstkowej $r_{jl\bullet\Omega}$ przyjmują wartości z przedziału $[-1, 1]$, a ich wartość może być większa lub mniejsza od wartości współczynnika korelacji całkowitej dla pary badanych cech. Ponadto wartości tych współczynników mogą się różnić znakiem. Interpretacja współczynnika korelacji cząstkowej dla dwóch badanych cech jest podobna do współczynnika korelacji liniowej Pearsona, z założeniem o wyłączeniu wpływu innych cech.

Współczynnik korelacji wielorakiej obliczamy na podstawie wzoru:

$$R_{j\bullet\Omega} = \sqrt{1 - \frac{\det(R)}{\det(R_{jj})}},$$

gdzie j w indeksie $j \bullet \Omega$ oznacza, że obliczamy współczynnik korelacji wielorakiej między cechą X_j a zespołem cech o numerach ze zbioru $\Omega = \{1, 2, \ldots, j-1, j+1, \ldots, m\}$, natomiast R_{jj} jest dopełnieniem algebraicznym macierzy korelacji R.

Współczynnik korelacji wielorakiej $R_{j\bullet\Omega}$ wyraża siłę korelacji i przyjmuje wartości z przedziału $[0, 1]$. Im bliższa jedności jest wartość współczynnika $R_{j\bullet\Omega}$, tym związek między cechą X_j a zespołem pozostałych cech jest silniejszy, natomiast im bardziej wartość tego współczynnika jest bliższa zeru, tym związek jest słabszy.
PRZYKŁAD.