PODSTAWY ANALIZY DANYCH POMIAROWYCH CZ. 4

Istotność parametrów dopasowania

Ważnym elementem analizy dopasowanego modelu jest sprawdzanie jak istotne są parametry, którymi jest opisany. Wyobraźmy sobie doświadczenie, którego wyniki są modelowo opisywane zależnością liniową $y = Ax$. W praktyce, bardzo często badacz nie wie jaki model opisuje jego dane, a jedynie go postuluje. Może się więc zdarzyć, że postulowany przez badacza model będzie postaci $y = Ax + B$. Model ten będzie również zgodny z danymi ale spodziewamy się, że otrzymane oszacowanie parametru B będzie bliskie zeru. Dlatego po każdym dopasowaniu, należy przeanalizować otrzymane oszacowania parametrów pod kątem ich istotności w zastosowanym modelu. Do tego celu stosuje się najprostsze testy (test 3σ, test t-Studenta), które w szybki sposób porównają otrzymane oszacowanie parametru (wraz z jego niepewnością) z zerem i odpowiedzą na pytanie jak prawdopodobne jest, że parametr ten jest istotny w dopasowaniu.

Ważne jest podkreślenia, że należy zachować szczególną ostrożność przy określaniu ilości parametrów. W szczególności należy unikać przeparametryzowania modelu, nieistotność parametru nie znaczy jednak automatycznie, że zmienną należy usunąć z modelu. W praktyce przy doborze parametrów objaśniających model należy zawsze kierować się zdrowym rozsądkiem i teorią dotyczącą badanego zagadnienia.

Badanie zależności parametrów dopasowania

Przypominamy, że w przypadku dopasowania modelu postaci $y(x, \mathbf{B})$ metodą najmniejszych kwadratów konstruujemy funkcję

$$\chi^2 = \sum_{i=1}^{N} \frac{(y_i - y(x_i, \mathbf{B}))^2}{u_i^2}$$

a następnie minimalizujemy ją względem wszystkich parametrów modelu $\frac{\partial \chi^2}{\partial B_i} = 0$, wyznaczając w ten sposób ocenę parametrów modelu $\widehat{\mathbf{B}}$.

Jak łatwo sobie wyobrazić w przypadku dopasowania modelu z więcej niż jednym parametrem pomiędzy poszczególnymi parametrami mogą występować zależności. Aby je badać oblicza się macierz kowariancji parametrów dopasowania (**B**), która stanowi uogólnienie pojęcia wariancji i kowariancji na przypadek wielowymiarowy. Macierz ta ma postać:

$$\Sigma(\mathbf{B}) = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1n} \\ \sigma_{21} & \sigma_2^2 & \cdots & \sigma_{2n} \\ \vdots & \cdots & \ddots & \vdots \\ \sigma_{n1} & \sigma_{n2} & \cdots & \sigma_n^2 \end{bmatrix}, \tag{1}$$

gdzie σ_i^2 jest wariancją i-tego parametru, a σ_{ij} kowariancją między i-tym i j-tym parametrem. Warto zauważyć, że macierz Σ jest macierzą symetryczną, której wyznacznik jest nieujemny.

Tak jak w przypadku analizy zależności dwóch zmiennych losowych, często wygodniejsza jest analiza macierzy korelacji. Macierz ta ma postać:

$$corr(\mathbf{B}) = \begin{bmatrix} 1 & \frac{\sigma_{12}}{\sigma_1\sigma_2} & \cdots & \frac{\sigma_{1n}}{\sigma_1\sigma_n} \\ \frac{\sigma_{21}}{\sigma_2\sigma_1} & 1 & \cdots & \frac{\sigma_{2n}}{\sigma_2\sigma_n} \\ \vdots & \cdots & \ddots & \vdots \\ \frac{\sigma_{n1}}{\sigma_n\sigma_1} & \frac{\sigma_{n2}}{\sigma_n\sigma_2} & \cdots & 1 \end{bmatrix}. \tag{2}$$

Wówczas, wartości wszystkich elementów macierzy należą do przedziału $\langle -1,1 \rangle$ (ponieważ są współczynnikami korelacji), wszystkie elementy leżące na diagonali macierzy równe są 1 (określa to stopień skorelowania i-tego parametru z nim samym), a wyznacznik tej macierzy należy do przedziału $\langle 0,1\rangle$. Wyznacznik macierzy korelacji jest miernikiem współliniowości (skorelowania) zmiennych objaśniających. Im bliższy 1, tym niższy jest stopień wzajemnego skorelowania zmiennych objaśniających. Im bliższy 0, tym siła tej korelacji większa. Wartości wyznacznika bliskie 0 świadczą o złym doborze parametrów objaśniających. Jeśli wyznacznik jest zbyt mały, musimy modyfikować model - powinniśmy z niego eliminować zmienne współliniowe. Istnienie silnej korelacji między zmiennymi objaśniającymi negatywnie wpływa na efektywność estymatorów parametrów modelu.

Jako ocenę wariancji parametru B_i przyjmujemy jego kwadrat niepewności. W celu obliczenia tej niepewności przyjmujemy, że wynika ona tylko z niepewności zmiennej y i korzystamy ze wzoru na błąd pośredni funkcji $B_i(y_i)$

$$u_{B_i}^2 = \sum_{i=1}^{N} u_i^2 \left(\frac{\partial B_i}{\partial y_i}\right)^2. \tag{3}$$

Jako ocenę kowariancji dwóch parametrów przyjmujemy

$$s_{B_i B_j} = \sum_{i=1}^{N} u_i^2 \left(\frac{\partial B_i}{\partial y_i}\right)\left(\frac{\partial B_j}{\partial y_i}\right), \tag{4}$$

a ich współczynnika korelacji:

$$r_{ab} = \frac{s_{B_i B_j}}{u_{B_i} u_{B_j}}. \tag{5}$$