

Kryteria wyboru adekwatnej funkcji trwałości ostrza skrawającego w programach krokowej regresji wielokrotnej

Cz. I - Metodyka

dr inż. Ryszard Filipowski

Instytut Techniki Wytwarzania, Wydział Inżynierii Produkcji, Politechnika Warszawska

Metoda krokowej regresji wielokrotnej pozwala wybrać funkcje o możliwie małej liczbie zmiennych niezależnych oraz ich interakcji. Opisano kryteria progowe selekcji zmiennych w programie krokowej regresji wielokrotnej *REGSTEP*, opracowanym na podstawie pracy [4] oraz w znanym programie *STATISTICA*

Wstęp

Funkcja regresji ustalona na podstawie wyników badań uzyskanych w/g przyjętego planu doświadczeń zawiera liniowe zmienne niezależne i ich interakcje w skali równomiernej lub w skali logarytmicznej. Funkcja regresji trzech zmiennych niezależnych u_1, u_2, u_3 z interakcjami $(u_1 \cdot u_2), (u_1 \cdot u_3), (u_2 \cdot u_3)$ w skali równomiernej ma postać:

$$\hat{z} = c_1 + c_2 \cdot u_1 + c_3 \cdot u_2 + c_4 \cdot u_3 + c_5 \cdot u_1 \cdot u_2 + c_6 \cdot u_1 \cdot u_3 + c_7 \cdot u_2 \cdot u_3 \quad (1)$$

lub w zapisie macierzowym:

$$z = U \cdot c \quad (2)$$

gdzie:

U – jest macierzą wejść uzupełnioną o pierwszą kolumnę zawierającą jedynki,

z – jest wektorem wyjść,

c – wektor współczynników funkcji regresji.

Funkcja regresji trzech zmiennych niezależnych $\ln(u_1), \ln(u_2), \ln(u_3)$ z interakcjami $[\ln(u_1) \cdot \ln(u_2)], [\ln(u_1) \cdot \ln(u_3)], [\ln(u_2) \cdot \ln(u_3)]$, w skali logarytmicznej ma postać:

$$\ln \hat{z} = \ln c_1 + c_2 \cdot \ln u_1 + c_3 \cdot \ln u_2 + c_4 \cdot \ln u_3 + c_5 \cdot \ln u_1 \cdot \ln u_2 + c_6 \cdot \ln u_1 \cdot \ln u_3 + c_7 \cdot \ln u_2 \cdot \ln u_3 \quad (3)$$

Funkcja regresji trzech zmiennych niezależnych $\ln(u_1), \ln(u_2), \ln(u_3)$ bez interakcji w skali logarytmicznej po przekształcaniu do postaci iloczynowej ma postać:

$$\hat{z} = e^{\ln c_1} \cdot u_1^{c_2} \cdot u_2^{c_3} \cdot u_3^{c_4}$$

Współczynniki c_i funkcji regresji (1), (3) są ustalane z warunku minimalizacji kwadratu odchyłek od hiperpłaszczyzny regresji [1]. Wektor współczynników regresji oblicza się ze wzoru [2]:

$$c = (U^T \cdot U)^{-1} \cdot U^T \cdot z \quad (4)$$

Do oceny istotności jednomianów w funkcji regresji w programie *REGSTEP* [4] stosuje się **kryterium proporcji** RE_{u_i} , które porównuje się z graniczną wartością zadawaną przez użytkownika programu (parametr *PCT*). Kryterium to akceptuje lub eliminuje jednomian w funkcji regresji. Wzór analityczny kryterium proporcji ma postać:

$$RE_{u_i} = \frac{\text{cov}(u_i, z)^2}{D^2(u_i - E(u_i)) \cdot D^2(z - E(z))} \quad (5)$$



LUBRICANTS.
TECHNOLOGY.
PEOPLE.



Środki smarne dla przemysłu

www.fuchs-oil.pl

Co nas wyróżnia:

- oferujemy najbardziej **innowacyjne** środki smarne, wszystkie nasze działania i projekty badawcze koncentrujemy na ich rozwoju,
- opracowujemy środki smarne **dopasowane** do wymogów i procesów produkcyjnych naszych Klientów,
- tworzymy **indywidualne** rozwiązania dla konkretnych zastosowań,
- dbamy, aby nasze produkty były **przyjazne środowisku** naturalnemu,
- naszym Klientom pomagamy wdrażać **oszczędne** rozwiązania poprzez optymalizację gospodarki smarowniczej.

Kompleksowa oferta:

- analizy serwisowe środków smarnych przeprowadzane w Laboratorium Kontroli Jakości Fuchs,
- stała obsługa logistyczno-doradcza Klienta,
- specjalistyczne doradztwo, wdrożenia i nadzór projektów technicznych,
- usługa outsourcingu gospodarki olejowej,
- budowa baz danych i nadzór nad dokumentacją,
- organizacja szkoleń i branżowych seminariów dla Klientów.



Do oceny istotności jednomianów w funkcji regresji w programie **STATISTICA** [5] stosuje się **kryterium testu F Snedecora**, które porównuje się z granicznymi wartościami tego testu wprowadzonymi przez użytkownika, jako kryterium $F_{\text{do wprowadzenia}}$ jednomianu lub $F_{\text{do usunięcia}}$ jednomianu z funkcji regresji. Wzór analityczny testu F Snedecora ma postać [2]:

$$F = \frac{\frac{1}{k} \sum_{i=1}^n (\hat{z}_i - \bar{z})^2}{\frac{1}{n-k-1} \sum_{i=1}^n (z_i - \hat{z}_i)^2} \quad (6)$$

gdzie:

\hat{z}_i - jest wartością funkcji regresji dla parametrów oznaczonych przez indeks i .

k - liczba współczynników w funkcji regresji.

$$(U^T \cdot U) = \begin{bmatrix} \sum_{i=1}^{i=k} (u_{i1} - \bar{u}_1)^2 & \sum_{i=1}^{i=k} (u_{i1} - \bar{u}_1) \cdot (u_{i2} - \bar{u}_2) & \dots & \sum_{i=1}^{i=k} (u_{i1} - \bar{u}_1) \cdot (u_{ik} - \bar{u}_k) \\ \sum_{i=1}^{i=k} (u_{i2} - \bar{u}_2) \cdot (u_{i1} - \bar{u}_1) & \sum_{i=1}^n (u_{i2} - \bar{u}_2)^2 & \dots & \sum_{i=1}^{i=k} (u_{i2} - \bar{u}_2) \cdot (u_{ik} - \bar{u}_k) \\ \dots & \dots & \dots & \dots \\ \sum_{i=1}^{i=k} (u_{ik} - \bar{u}_k) \cdot (u_{i1} - \bar{u}_1) & \sum_{i=1}^{i=k} (u_{ik} - \bar{u}_k) \cdot (u_{i2} - \bar{u}_2) & \dots & \sum_{i=1}^{i=k} (u_{ik} - \bar{u}_k)^2 \end{bmatrix} \quad (10)$$

Test **F Snedecora** oblicza się w obu programach dla kolejno ustalanych funkcji regresji \hat{z} , w których zmienia się liczba zmiennych niezależnych. Dla każdego kroku regresji program oblicza oprócz **testu F** pozostałe testy statystyczne, tj. współczynnik **korelacji wielokrotnej R**, oraz testy **t Studenta** dla ustalonych współczynników w funkcji regresji. Wzór analityczny współczynnika korelacji wielokrotnej **R** oblicza się w programie wg wzoru [2]:

$$R^2 = \frac{\sum_{i=1}^n (\hat{z}_i - \bar{z})^2 + \sum_{i=1}^n (z_i - \hat{z}_i)^2}{\sum_{i=1}^n (z_i - \hat{z}_i)^2} \quad (7)$$

gdzie:

$\sum_{i=1}^n (z_i - \hat{z}_i)^2$ = wariancja resztowa (lub wariancja nie wyjaśniona przez model [3]).

Test **t Studenta** dla współczynników funkcji regresji c_i oblicza się w programach wg wzoru [2]:

$$t = \frac{c_i}{\sqrt{\frac{1}{k} (z_i - \hat{z}_i)^2 \cdot \sqrt{D^2(z_i - E(z_i))}}} \quad (8)$$

Po ostatecznym ustaleniu funkcji regresji przez program **REGSTEP** lub **STATISTICA** jego istotność określa użytkownik poprzez analizę statystyczną wielkości obliczanych przez te programy, tj.: ocenę współczynnika **korelacji wielokrotnej R**, **testu F Snedecora** oraz **t Studenta** dla współczynników w funkcji regresji oraz dodatkowo, poprzez **analizę reszt** (ang. *residuals*).

Kryterium progowe proporcji wariancji wprowadzenia zmiennej do funkcji regresji w programie krokowej regresji REGSTEP

Kryterium proporcji wariancji oblicza się w programie krokowej regresji **REGSTEP** [4] za pomocą macierzy kowariancji ($U^T \cdot U$) zmiennych wejściowych u_i względem ich wartości średnich \bar{u}_i .

Macierz danych wejściowych U ma postać:

$$U = \begin{bmatrix} (u_{11} - \bar{u}_1) & (u_{12} - \bar{u}_2) & \dots & (u_{1k} - \bar{u}_k) \\ (u_{21} - \bar{u}_1) & (u_{22} - \bar{u}_2) & \dots & (u_{2k} - \bar{u}_k) \\ \dots & \dots & \dots & \dots \\ (u_{i1} - \bar{u}_1) & (u_{i2} - \bar{u}_2) & \dots & (u_{ik} - \bar{u}_k) \end{bmatrix} \quad (9)$$

Macierz kowariancji ($U^T \cdot U$) przyjmuje postać:

Macierz kowariancji jest sumą iloczynów krosowych odchyłek zmiennych od ich wartości średnich. Elementami leżącymi na przekątnej macierzy ($U^T \cdot U$) są w kolejności wariancje odchyłek zmiennych od ich wartości średnich, tj.:

$$\sum_{i=1}^n (u_{i1} - \bar{u}_1)^2, \dots, \sum_{i=1}^n (u_{i2} - \bar{u}_2)^2, \dots, \sum_{i=1}^n (u_{ik} - \bar{u}_k)^2 \quad (11)$$

Ostatnia wariancja o indeksie k dotyczy zmiennej zależnej.

Macierz kowariancji wykorzystuje się w programie **REGSTEP** do obliczenia **kryterium proporcji** RE_{u_i} w celu wprowadzenia zmiennych do funkcji regresji.

Kryterium proporcji wariancji stosowane do wprowadzenia zmiennych do funkcji regresji, przyjmuje postać:

$$RE_{u_i} = \frac{\left[\sum_{i=1}^n (u_{ij} - \bar{u}_j) \cdot (u_{ik} - \bar{u}_k) \right]^2}{\sum_{i=1}^n (u_{ij} - \bar{u}_j)^2 \cdot \sum_{i=1}^n (u_{ik} - \bar{u}_k)^2} \quad (12)$$

Dla demonstracji obliczeń **kryterium** RE_{u_i} poniżej **przedstawiono** macierz kowariancji dla trzech zmiennych niezależnych i czwartej zależnej **w skali równomiernej** dla stali C45 (13). W ostatniej kolumnie macierzy podane są kowariancje zmiennych względem zmiennej zależnej.

$$RE_{u_i} = \frac{\left[\sum_{i=1}^n (u_{ij} - \bar{u}_j) \cdot (u_{ik} - \bar{u}_k) \right]^2}{\sum_{i=1}^n (u_{ij} - \bar{u}_j)^2 \cdot \sum_{i=1}^n (u_{ik} - \bar{u}_k)^2} \quad (13)$$

Na przykład, dla zmiennej u_1 , kryterium proporcji RE_{u_1} , w skali równomiernej (12) wynosi:

$$RE_{u_1} = \frac{(-18850,61)^2}{39372,29 \cdot 18709,09} = 0,48$$

W podobny sposób obliczane są wartości kryterium proporcji RE_{u_i} dla pozostałych zmiennych niezależnych u_i , $i = 2, \dots, k$. Obliczone kryteria proporcji RE_{u_i} dla zmiennych u_i oraz ich interakcji zestawia się z założoną progową wartością PCT . Jeśli różnica dla analizowanej zmiennej ($RE_{u_i} - PCT$) ≥ 0 , wtedy zmienna jest dodawana do modelu regresji. Procedura ta jest wykonywana do momentu uzyskania równania ze zmiennymi mającymi istotny wpływ na zmienną zależną \hat{z} .

Kryterium progowe testu F wprowadzenia zmiennej do funkcji regresji w programie krokowej regresji STATISTICA

Test F jest stosunkiem, dwóch wariancji z uwzględnieniem ich stopni swobody (6) [2]. Funkcja regresji \hat{z} występująca we wzorze (6) jest modyfikowana w kolejnych krokach ustalania postaci ostatecznej funkcji regresji w programie STATISTICA [5] podczas sprawdzania istotności rozważonego jednomianu. Metoda sprawdzania istotności jednomianu (przytoczona w [3]) jest następująca. Ustala się standardowo krytyczną wartość testu $F_{\text{do wprowadzenia}} = 1$ oraz $F_{\text{do usunięcia}} = 0$. W kolejnych krokach określa się wartość zmienionej funkcji regresji przez dodanie kolejnych jednomianów. W pierwszym kroku funkcja regresji ma postać:

$$\hat{z} = c_0 + c_1 \cdot u_j, \quad j = 1, 2, \dots, k \quad (14)$$

k – oznacza liczbę zmiennych rozważanych w planie badań.

Istotność zmiennej oznaczonej indeksem j weryfikuje się w/g hipotezy zerowej: $H_0: c_1 = 0$, przy hipotezie alternatywnej $H_1: c_1 \neq 0$. Zmienna kandydująca o największej wartości testu F wprowadzana jest do funkcji regresji jeśli spełnia założoną wartość progową tego testu. W następnym kroku budowany jest model w postaci:

$$\hat{z} = -c_0 + c_1 \cdot u_1 + c_2 \cdot u_j, \quad j = 2, \dots, k \quad (15)$$

Do modelu dodaje się tę zmienną, dla której wartość *statystki F* większa od progowej jest największa. Podobnie postępuje się z pozostałymi zmiennymi, ich interakcjami i potęgami aż do momentu uzyskania funkcji regresji ze zmiennymi, które mają istotny wpływ na zmienną zależną \hat{z} .

Uzyskana funkcja regresji w/g krokowej regresji i weryfikowana statystycznie poprzez wartości testów: RE , F , R , t *Studenta*, winna być dodatkowo sprawdzana przez *analizę reszt* tj. różnicę wartości wyjściowej i wartości funkcji regresji. Ostateczną decyzję przyjęcia postaci funkcji regresji winien podjąć prowadzący badania, gdyż automatyczny dobór funkcji przez program komputerowy może być czasami zawodny.

Wnioski

Funkcja regresji może zawierać zmienne, ich interakcje oraz potęgi zmiennych, tj. postać bardziej złożoną niż przedstawioną we wzorach (1) i (3). Przy dużej liczbie zmiennych niezależnych zaleca się stosować do ustalenia postaci funkcji regresji program krokowej regresji wielokrotnej. Przy przyjętych wartościach progowych dla wprowadzania zmiennych do funkcji regresji, program krokowej regresji wielokrotnej eliminuje niektóre interakcje lub potęgi zmiennych, co pozwala uzyskać bardziej zwartą i zarazem dokładną (tj. adekwatną) postać tej funkcji.

Literatura

- [1] Draper N.R., Smith H.: Applied Regression Analysis, second edition. John Wiley & Sons, New York 1981.
- [2] Kukielka L.: Podstawy badań inżynierskich, PWN, Warszawa 2002,
- [3] Stanisław A.: Przystępny Kurs Statystyki, Tom II, Kraków 2007, StatSoft.
- [4] System/360 Scientific, Subroutine Package (360 A – CM 03X) Version III, IBM Technical Publication Department, Fifth Edition 1970.
- [5] Program STATISTICA, www.StatSoft.pl

reklama



obróbka metalu
kwartalnik techniczny

Fachowe i rzetelne źródło informacji

ZAMÓW PRENUMERATĘ
www.e-obrobkametalu.pl/pl/prenumerata