FOLOSIREA REGRESIEI LINIARE MULTIPLE IN ANALIZA REZULTATELOR EXPERIMENTARILOR

HEREA-BUZATU Constantin, sef lucr.dr.ing., Universitatea Tehnica de Constructii Bucuresti

This paper shows how to make the analysis of the results obtained from the experiments.

Regresia liniara multiplă analizează influenţa a *n* variabile independente $x_1, x_2, \ldots, x_n$ asupra unei variabile independente *y*. Considerând că în cadrul experimentului se fac *n* determinări rezultă că încercarea *i* este caracterizată de punctul (x_{1i} , x_{2i} , ...x_{ri} ,y_i), unde i = 1, 2,...,*n*.

Presupunând că între variabilele independente şi cea dependentă există o relaţie liniară, media distribuţiei lui *y* este de forma:

$$\mu_y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_r x_r \tag{1}$$

Aplicând metoda celor mai mici pătrate obţinem estimaţii ale coeficienţilor β_0 , β_1 ,…β_r. Pentru aceasta vom pune condiţia:

$$\sum_{i=1}^{n} [y_i - (\beta_0 + \beta_1 x_{1i} + \beta_2 x_{2i} + \ldots \beta_r x_{ri})]^2 = \min \tag{2}$$

Derivând (2) în raport cu β_0 , β_1 ,…β_r şi notând cu b_0, b_1,…b_r estimaţiile coeficienţilor de regresie, derivatele parţiale constitue un sistem de ecuaţii normale:

$$\begin{aligned}
&\sum_{i=1}^{n} y_i = nb_0 + b_1 \sum_{i=1}^{n} x_{1i} + b_2 \sum_{i=1}^{n} x_{2i} + \ldots b_r \sum_{i=1}^{n} x_{ri} \\
&\sum_{i=1}^{n} y_i = b_0 \sum_{i=1}^{n} x_{1i} + b_1 \sum_{i=1}^{n} x_{1i}^2 + b_2 \sum_{i=1}^{n} x_{1i} x_{2i} + \ldots + b_r \sum_{i=1}^{n} x_{1i} x_{ri} \\
&\sum_{i=1}^{n} y_i = b_0 \sum_{i=1}^{n} x_{2i} + b_1 \sum_{i=1}^{n} x_{2i} x_{1i} + b_2 \sum_{i=1}^{n} x_{2i}^2 + \ldots + b_r \sum_{i=1}^{n} x_{2i} x_{ri} \\
&\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\
&\sum_{i=1}^{n} y_i = b_0 \sum_{i=1}^{n} x_{ri} + b_1 \sum_{i=1}^{n} x_{ri} x_{1i} + b_2 \sum_{i=1}^{n} x_{ri} x_{2i} + \ldots + b_r \sum_{i=1}^{n} x_{ri}^2
\end{aligned} \tag{3}$$

Având în vedere cele arătate în [12], în continuare este analizată, printr-o regresie multiplă, influenţa factorilor aleşi asupra presiunii şi a eroziunii.

Această analiză se realizează cu ajutorul programului *Microsoft Office Excel,* folosind funcţia LINEST.

Această funcţie calculează statisticile pentru o dreaptă prin metoda celor mai mici pătrate, dreaptă estimată a se apropia cel mai mult de perechile de puncte obţinute experimental.

Ecuaţia dreptei estimate este

$$y = m_1 x_1 + m_2 x_2 + ... + m_n x_n + b \tag{4}$$

unde valoarea dependentă *y* este o funcţie de valori independente *x*. Valorile *m* sunt coeficienţi corepunzători fiecărei valori *x,* iar *b* este o valoare constantă. Este de remarcat că *y*, *x* şi *m* pot fi nişte vectori.

Matricea formată cu funcţia LINEST este de forma:

$$\{m_n \cdot m_{n-1} \cdot ... m_1 \cdot b\} \tag{5}$$

Funcţia LINEST poate, de asemenea, să determine, statistici de regresie suplimentară, matricea devenind:

$$\{m_n \cdot m_{n-1} ... m_1 \cdot b; se_{n-1} \cdot se_{n-1} \cdot ... se_1 \cdot se_b; r2 \cdot se_y; F \cdot df; ss_{reg} \cdot ss_{resid}\} \tag{6}$$

Statisticile de regresie suplimentare sunt următoarele:

se_1, se_2, … se_n – valorile erorii tip corespunzătoare coeficienţilor m_1, m_2, … m_n.

se_b – valoarea erorii tip corespunzătoare constantei *b*;

*r*2– coeficientul de determinare. Compară valorile *y* estimate cu valorile *y* reale şi variază între 0 şi 1. Un coeficient de determinare egal 1 indică o corelare perfectă (nici o diferenţă) între valorile *y* estimate şi valorile *y* reale, determinate experimental. Invers un coeficient de determinare egal cu zero arată că ecuaţia de regresie nu poate servi la estimarea unei valori *y* ;

se_y – eroarea tip pentru valoarea *y* estimată;

F – statistica *F* sau valoarea *F* observată. Acest parametru se foloseşte pentru a constata dacă relaţia observată între variabila dependentă şi variabilele independente este datorată întâmplării;

df – gradele de libertate. Ele ajută la găsirea valorilor critice ale statisticii *F*. Aceste valori critice se compară cu valoarea *F* determinată prin funcţia LINEST, pentru a evalua nivelul de încredere al modelului utilizat în experimentări;

ss_{reg} – suma de regresie a pătratelor;

ss_{resid} – suma reziduală a pătratelor.

Ordinea în care statisticile de regresie suplimentare sunt determinate este ilustrată în tabelul următor:

	A	B	C	D	E	F
1	m_n	m_{n-1}	…	m_2	m_1	b
2	se_n	se_{n-1}	…	se_2	se_1	se_b
3	$r2$	se_y				
4	F	df				
5	ss_{reg}	ss_{resid}				

Orice dreaptă poate fi descrisă prin panta sa şi prin ordonata sa la origine. Pentru a determină panta unei drepte, notată în general cu litera *m*, se iau două puncte ale dreptei (x_1, y_1) şi (x_2, y_2). Panta va fi egală cu :

$$m = \frac{y_2 - y_1}{x_2 - x_1} \quad (7)$$

Ordonata la origine, notată în general cu litera *b*, este valoarea lui *y* a punctului de intersecţie a dreptei cu axa ordonatelor (*y*).

Ecuaţia unei drepte este:

$$y = mx + b \quad (8)$$

Odată cunoscute valorile lui *m* şi *b*, fiecare punct al dreptei poate fi calculat atribuind o valoare lui *x* sau *y* în ecuaţia dreptei.

În cazul în care nu se dispune decât de o singură variabilă independentă *x*, calculele lui *m* şi *b* se bazează pe următoarele formule:

$$m = \frac{\Sigma(x - \bar{x})(y - \bar{y})}{\Sigma(x - \bar{x})^2}, \quad b = \bar{y} - m\bar{x} \quad (9)$$

unde: $\bar{x}$ şi $\bar{y}$ sunt mediile valorilor determinărilor.

Într-o analiză a regresiei, Microsoft Excel calculează, pentru fiecare punct pătratul diferenţei dintre valorile *y* estimate şi valorile *y* reale suma acestor diferenţe la pătrat (pătratice) se numeşte *„sumă reziduală de pătrate” – ss_{resid}*.

Microsoft Excel calculează după aceea *suma totală a pătratelor – ss_{total}.*

Dacă argumentul constantă este ADEVĂRAT sau neglijabil, suma totală de pătrate este *suma de diferenţe la pătrat între valorile y reale şi media valorilor y.*

Dacă argumentul este FALS, suma totală a pătratelor este egală cu suma pătratelor valorilor *y* reale (fără a scădea valoarea medie *y* din fiecare valoare individuală).

După aceea *suma regresiei pătratelor, ss_{reg}*, poate fi găsită făcând:

$$ss_{reg} = ss_{total} - ss_{resid} \quad (10)$$

Cu cât suma reziduală a pătratelor este mai mică în comparaţie cu suma totală a pătratelor şi cu cât valoarea coeficientului de determinare r2 este mai mare (mai aproape de valoarea 1), cu atât ecuaţia rezultantă a analizei de regresie explică releţia între variabile de o manieră mai satisfăcătoare.

Coeficientul *r2* este egal cu:

$$r2 = \frac{ss_{reg}}{ss_{total}} \quad (11)$$

În anumite cazuri, una sau mai multe coloane *x* (presupunând că valorile *y* şi *x* sunt scrise în coloane) pot să nu aibă nici o valoare *previzibilă suplimentară* în prezenţa altor coloane *x*. Cu alte cuvinte, eliminarea uneia sau a mai multor coloane *x* poate conduce la valori *y* prestabilite la fel de precise. În acest caz, coloanele *x* de prisos trebuie neglijate din modelul regresiei. Acest fenomen se numeşte *coliniaritate*, pentru că nu are importanţă care coloană de prisos, poate fi exprimată ca o sumă de multipli de coloane *x* utile.

Funcţia LINEST verifică coliniaritatea şi suprimă toate coloanele *x* de prisos descoperite în modelul regresiei. *Coloanele x suprimate se disting în rezultatele funcţiei LINEST prin faptul că ele au valoarea zero pentru coeficienţii m şi se (valoarea erorii tip corespunzătoare lui m_n, b, y).*

Dacă una sau mai multe coloane sunt suprimate ca fiind neglijabile, statisticile *df* sunt influenţate, deoarece ele depind de numărul de coloane *x* utilizate cu finaluri previzibile.

Dacă valoarea lui *df* este modificată din cauza eliminării coloanelor *x* neglijabile, valorile lui se_y şi *F* sunt de asemenea influenţate.

Valoarea lui df este calculată ca şi cum nici o coloană *x* n-ar fi eliminată din model din cauza coliniarităţii.

Dacă sunt *k* coloane cu valori *x* cunoscute şi dacă argumentul constantă este ADEVĂRAT sau neglijabil, atunci:

$$df = n - k - 1 \tag{12}$$

Dacă argumentul constantă este FALS, atunci:

$$df = n - k \tag{13}$$

În cele două cazuri, fiecare coloană diminuată din cauza coliniarităţii creşte valoarea lui *df* cu 1.

Este de remarcat că valorile *y* presupuse prin ecuaţia de regresie pot să nu fie validate dacă ele se găsesc în afara domeniului de valori *y* utilizate pentru determinarea acestei ecuaţii.

*Valoarea ridicată a lui r*2 (apropiată de 1) lasă să se presupună o relaţie strânsă între variabilele independente *x* şi valoarea dependentă *y*.

Statistica F ne permite să apreciem dacă rezultatele reprezentând valoarea ridicată a lui *r*2 sunt rodul întâmplării.

Pentru început se presupune că nu există nici o relaţie adevărată între variabilele *x* şi *y* , deşi analiza statistică a datelor experimentale demonstrează o relaţie strânsă. *Se numeşte Alpha probabilitatea de a se înşela concluzionând existenţa unei relaţii*.

Valorile statisticii *F* obţinute prin intermediul funcţiei LINEST sunt comparate cu valori critice ale sale, valori fie luate din tabele de distribuţie, fie determinate cu ajutorul unei alte funcţii din Microsoft Excel, FDIST.

Distribuţia *F* corespunzătoare are gradele de libertate v_1 şi v_2. Dacă *n* este numărul de determinări experimentale şi argumentul constantă este ADEVĂRAT sau neglijabil, atunci:

$$v_1 = n - df - 1 \quad \text{şi}$$

$$v_2 = df \tag{14}$$

Dacă argumentul constantă este FALS, atunci:

$$v_1 = n - df \quad \text{şi}$$

$$v_2 = df$$

Funcţia Excel, FDIST (*F*, v_1, v_2) determină probabilitatea ca o valoare *F*, obţinută în funcţia LINEST, mai mare decât cea critică, să fie datorată întâmplării.

Dacă această probabilitate este mică, putem concluziona că ecuaţia de regresie este utilă în stabilirea unei relaţii între variabilele independente *x* şi valoarea dependentă *y*, aşa cum statistica *r*2 arătase deja.

Calculul unei alte statistici, notată cu T, permite să se determine dacă fiecare coeficient al pantei dreptei intervine în estimarea valorii *y*.

Pentru aceasta se împarte valoarea coeficientului pantei m_n la valoarea erorii tip estimată se_n, (valori corespunzătoare variabilei x_n). Aceasta dă valoarea observată t_n

$$t_n = \frac{m_n}{se_n} \tag{15}$$

Dacă valoarea absolută a lui *t* este suficient de mare, în raport cu o valoare critică a sa, se poate concluziona că **coeficientul pantă este util în estimarea valorii *y***.

Valorile critice pentru *t* pot fi găsite în tabele statistice publicate sau pot fi calculate cu ajutorul altei funcţii din Microsoft Excel, TINV (Alpha, *df*).

În măsura în care valoarea absolută a lui t este superioară valorii critice, variabila corespunzătoare x este o variabilă semnificativă în estimarea valorii dependente y. Se poate astfel testa semnificatia statistică a fiecărei dintre variabilele independente x.

BIBLIOGRAFIE

[1] HEREA-BUZATU ,C. – *Consideratii asupra determinarii experimentale a influentei adancimii in abraziv asupra presiunii presiunii pe suprafata palelor frezelor dragi;or refulante in procesele de lucru, ,* UTCB, SINUC, Bucureşti, 2006.

[2] HEREA-BUZATU,C. - *Unele rezultate experimentale privind influenta unghiului de atac asupra presiunii pe suprafata palelor frezelor dragilor refulante in procesul saparii pamanturilor,* UTCB, SINUC, Bucureşti, 2006.

[3] HEREA-BUZATU, C. – *Dispozitiv şi metoda de determinare experimentală a presiunii şi a uzurii prin eroziune abrazivă la palele frezelor drăgilor refulante,* UTCB, SINUC, Bucureşti, **2006**.

[4] HEREA-BUZATU, C. –*Determinarea experimentala a influientei vitezei abrazivului la diverse unghiuri de atac asupra presiunii pe suprafata palelor frezelor dragilor refulante in procesele de sapare***,** UTCB,SINUC,Bucuresti,2007.

[5**]** HEREA-BUZATU, C. –*Rezultate experimentale ale influientei dimensiunii particulelor si a unghiului de atac asupra presiunii pe suprafata palelor frezelor dragilor refulante in procesele de lucru,* UTCB,SINUC,Bucuresti,2007.

[6] HEREA-BUZATU, C. – *Determinarea experimentala a eroziunii palelor frzelor dragilor refulante in cazul nisipului..* UTCB,SINUC,Bucuresti,2008.

[7] HEREA-BUZATU, C. – *Determinarea experimentala a eroziunii palelor frzelor dragilor refulante in cazul pietrisului..* UTCB,SINUC,Bucuresti,2009.

[8] HEREA-BUZATU, C. – *Determinarea experimentala a eroziunii palelor frzelor dragilor refulante in cazul criblurii .* UTCB,SINUC,Bucuresti,2010.

[9] HEREA-BUZATU, C. – *Influenta presiunii,a formei si dimensiunii particulelor abrazive asupra uzurii palelor frezelor dragilor absorbant refulante.* UTCB,SINUC,Bucuresti,2011.

[10] HEREA-BUZATU,C.—*Influenta duratei de exploatare asupra uzurii palelor frezelor dragilor absorbant refulante.* UTCB,SINUC,Bucuresti,2012.

[11] CRETU,T.,FALIE,V.-- *Prelucrarea datelor experimentale in fizica*. ED. DIDACTICA SI PEDAGOGICA,Bucuresti,1987.

[12] HEREA-BUZATU,C –*Analiza rezultatelor experimentarilor cu privire la uzarea frezelor dragilor absorbant refulante*,UTCB,SINUC,Bucuresti.2013.