

# METODE DE PRELUCRARE A DATELOR EXPERIMENTALE

.....  
Membru corespondent al A.Ș.M.  
Evgheni Lvovschi

The following articles describe methods of processing mathematic-statistic experimental data, also with the help of the computer. The research works with the help of these methods will obtain the most information from the results of the fulfilled experiments.

## Partea I. Experimente pasive.

Noțiunea „experimente pasive” necesită comentarii. Această noțiune a apărut în secolul trecut pentru a face diferență între experimente obișnuite clasice, *pasive*, în care în cursul experimentării se variază numai un singur factor  $x$ , și noul tip de experimente, *active*, în care în timpul experimentării se variază mai mulți factori  $x_j$  conform unui plan special. Mulți savanți nu admit noțiunea „experimente pasive”. Ei susțin că nu există niciun experiment pasiv. În fiecare experiment, chipurile, se implică activ cercetătorul. Poate că așa și este. Însă aceasta este terminologia științifică, acceptată în toată lumea.

Pe parcursul multor ani autorul a examinat o problema foarte actuală pentru știință. Este vorba de elaborarea modelelor matematice optime pe baza datelor de observație pentru descrierea diferitelor fenomene. În urma acestor cercetări a fost obținută o nouă metodă complexă cu multe trepte pentru construcția modelelor, inclusiv cu evidența sensului fizic al fenomenelor studiate.

Se știe că numărul modelelor posibile pentru descrierea unui fenomen  $y$  depinde de numărul factorilor  $x$ , care sunt luați în evidență. De exemplu, în cel mai simplu caz, când  $y$  depinde numai de un factor  $x$ , modele posibile pot fi:

$$y = ax \quad (1)$$

$$y = a + bx \quad (2)$$

$$y = a + b/x \quad (3)$$

$$y = 1/(a + bx) \quad (4)$$

$$y = x/(a + bx) \quad (5)$$

$$y = ab^x \quad (6)$$

$$y = ae^{bx} \quad (7)$$

$$y = 1/(a + be^{-x}) \quad (8)$$

$$y = ax^b \quad (9)$$

$$y = a + b \lg x \quad (10)$$

$$y = a/(b + x) \quad (11)$$

$$y = ax/(b + x) \quad (12)$$

$$y = ae^{b/x} \quad (13)$$

$$y = a + bx^n \quad (14)$$

Așa formule pot fi multe, dar totuși numărul lor este în cel mai simplu caz, care se studiază, limitat. De observat că toate formulele acestea pot fi liniarizate și, ca rezultat, coeficienții  $a, b$  pot fi determinați pe baza datelor de observații cu metoda pătratelor minime. În cazul, în care numărul factorilor  $x_j$  este mai mare decât 1 ( $j > 1$ ), numărul tipurilor de modele crește la infinit [2].

Sistemul de ecuații, așa numite normale, poate fi liniar, liniarizabil și neliniar. În primele două cazuri pentru obținerea modelelor matematice pe baza datelor experimentale poate fi utilizată metoda pătratelor minime. Ultimul caz e mai complicat. Aici, pentru a determina coeficienții modelului matematic, este necesar de a rezolva un sistem de ecuații algebrice neliniare, ceea ce, după cum se știe, nu e deloc simplu. Metoda pătratelor minime este un caz particular a metodei de verosimilitate maximă, când datele de observație sunt distribuite normal (așa numita distribuția lui Hauss). Dacă funcția  $y$  (fenomenul examinat, variabila dependentă) este o valoare continuă (nu discretă), atunci probabilitatea faptului că aceasta funcție în populația statistică generală va avea o oarecare valoare este:

$$P(y, \beta)dy, \quad (15)$$

unde  $\beta$  este un parametru necunoscut, care va fi determinat în cursul rezolvării problemei.

Formând un eșantion finit cu volumul  $n$  din populația statistică generală, obținem valorile:

$$y_1, y_2, \dots, y_n \quad (16)$$

Procesul de sustragere a membrilor eșantionului trebuie să fie așa, ca probabilitatea obținerii noii valori să fie independentă de valorile obținute precedent. Probabilitatea de a obține valorile (15) este egală:

$$dP = Ldy_1 dy_2 \dots dy_n, \quad (17)$$

unde:

$$L(y_1, y_2, \dots, y_n; \beta) = P\{y_1, \beta\}P\{y_2, \beta\} \dots P\{y_n, \beta\}. \quad (18)$$

Funcția  $L$  poartă numele de funcție de verosimilitate. Estimarea în eșantion a parametrului  $\beta$  este parametrul  $b$ .

În practică este mai comod de a opera nu cu funcția  $L$ , ci cu logaritmul ei, de aceea determinarea parametrului  $b$  poate fi efectuată prin rezolvarea ecuației cu derivata parțială:

$$\frac{\partial \lg L(y; b)}{\partial b} = 0 \quad (19)$$

Dacă distribuția (repartiția) valorii  $y$  în populația statistică generală este **normală**, atunci probabilitatea de a obține  $y$  între  $y$  și  $y+dy$  este:

$$P\langle y; \eta; \sigma^2 \rangle dy = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(y-\eta)^2}{2\sigma^2}\right\} dy \quad (20)$$

Pentru un eșantion aleator cu volumul  $n$  se poate de scris:

$$L\langle y; \eta; \sigma^2 \rangle = \frac{1}{(2\pi\sigma^2)^{n/2}} \times \exp\left\{-\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \eta)^2\right\} \quad (21)$$

Este clar că funcția de verosimilitate  $L$  va atinge **maximumul** în cazul în care suma pătratelor devierilor  $\sum_{i=1}^n (y_i - \eta)^2$  va fi **minimă**. De aceea și metoda se numește metoda pătratelor minime.

Din aceasta se obține ecuația generală a metodei pătratelor minime:

$$U = \sum_{i=1}^n W_i \left[ y_i - \eta(\vec{\beta}, \vec{x}) \right]^2 \quad (22),$$

unde  $\eta(\vec{\beta}, \vec{x})$  este funcția de la variabila independentă de orice formă (liniară, neliniară) cu un sistem (vector) de coeficienți  $\vec{\beta}$ , constanți pentru fiecare fel de funcție, care de fiecare dată trebuie determinați.

În cel mai simplu caz, când sistemul de parametri este liniar, ecuația (22) se simplifică:

$$U = \sum_{i=1}^n \frac{1}{\sigma_{y_i}^2} \left[ y_i - (\beta_0 + \sum_{j=1}^p b_j x_{ij}) \right]^2 \quad (23)$$

în care  $1/\sigma_{y_i}^2 = W_i$  - coloniță-vector de greutate specifice,  
 $i$  - numărul încercărilor,  
 $j$  - numărul factorilor în regresie multiplă.

Așa dar, *metoda pătratelor minime este cazul particular al metodei de verosimilitate maximală, când distribuția funcției  $Y$  este normală.*

Pentru cel mai simplu caz, când funcția  $Y$  depinde numai de un factor  $X$ , are loc așa numita dependență-pereche și formula (23) se transformă și se simplifică:

$$U = \sum_{i=1}^n W_i [y_i - (\beta_0 + \beta_1 x_i)]^2 \quad (24)$$

Sistemul ecuațiilor normale a metodei pătratelor minime va fi obținut prin executarea operațiunii:

$$\frac{\partial U}{\partial b_0} = 0, \quad \frac{\partial U}{\partial b_1} = 0. \quad (25)$$

$$\sum_{i=1}^n W_i [y_i - (b_0 + b_1 x_i)] = 0 \quad (26)$$

$$\sum_{i=1}^n W_i [y_i - (b_0 + b_1 x_i)] \times x_i = 0$$

Prin rezolvarea acestor ecuații se determină coeficienții  $b_0$  și  $b_1$  a modelului liniar  $y = b_0 + b_1 x$ .

În folosirea metodei pătratelor minime se pot întâlni trei situații.

**Prima situație.** Fiecărei valori fixate a factorului  $x_i$  este atribuită numai o singură valoare a funcției  $y$ . Dispersia  $y$  nu există. Formula (26) se scrie:

$$\sum_{i=1}^n [y_i - (b_0 + b_1 x_i)] = 0 \quad (27)$$

$$\sum_{i=1}^n [y_i - (b_0 + b_1 x_i)] \times x_i = 0$$

**Situația a doua.** Fiecărei valori fixate a factorului  $x_i$  se atribuie  $k$  valori a variabilei dependente  $y$ , persistă distribuția normală, media teoretică  $\eta_i$  și dispersia  $\sigma_{y_i}^2$ . Pentru a rezolva această problemă se folosește sistemul de ecuații (26). Persistă membrul

$$W = \frac{1}{\sigma_{y_i}^2}.$$

**Situația a treia.** Aici factorul  $x$  aidoma funcției  $y$  sunt determinate cu erori. În loc de o valoare fixată a factorului  $x$ , în fiecare experiment sunt  $k_x$  valori ai factorului  $x$  cu distribuția normală, media teoretică  $\xi_i$  și dispersia  $\sigma_{x_i}^2$  și  $k_y$  valori pentru variabila dependentă  $y$  cu distribuția normală, media teoretică  $\eta_i$  și dispersia  $\sigma_{y_i}^2$ . Distribuția  $x$  nu este corelată cu distribuția  $y$ . Pentru estimarea coeficienților  $\beta_0$  și  $\beta_j$  în acest caz metoda pătratelor minime nu este eficientă și se aplică altă metodă, anume metoda **confluentă**. Metoda aceasta, fiind destul de complicată, se aplică de fapt numai în fizica nucleară și aici nu se examinează.

Așa dar, poate fi scrisă ecuația metodei pătratelor minime în cazul, când parametrizarea este liniară:

$$U = \sum_{i=1}^n \left[ y_i - \left( b_0 + \sum_{j=1}^k b_j x_{ij} \right) \right]^2 \quad (28),$$

apoi se execută operațiunea:

$$\frac{\partial U}{\partial b_0} = 0, \frac{\partial U}{\partial b_j} = 0 \quad (29)$$

și se obține sistemul de ecuații a metodei pătratelor minime în cazul, când persistă mai mulți factori  $x_j$  și parametrizarea este liniară. În forma de matrice acest sistem se scrie:

$$(X^* X)B = X^* Y \quad (30)$$

$$B = \begin{pmatrix} b_0 \\ b_1 \\ b_2 \\ \bullet \\ b_j \\ \bullet \\ b_p \end{pmatrix} \quad (31)$$

$X$  – matricea tuturor valorilor factorilor examinați:

$$X = \begin{pmatrix} x_{10} x_{11} x_{12} \dots x_{1j} \dots x_{1p} \\ x_{20} x_{21} x_{22} \dots x_{2j} \dots x_{2p} \\ \dots \\ x_{i0} x_{i1} x_{i2} \dots x_{ij} \dots x_{ip} \\ \dots \\ x_{n0} x_{n1} x_{n2} \dots x_{nj} \dots x_{np} \end{pmatrix} \quad (32)$$

$x_{i0}$  – vectorul de valori, care determină membrul liber al ecuației  $b_0$ . În matricea de date inițiale valorile sunt egale cu 1.

$X^*$  - matricea sistemului de ecuații normale:

$$X^* = \begin{pmatrix} n \sum x_{i1} \sum x_{i2} \dots \sum x_{ij} \dots \sum x_{ip} \\ \sum x_{i1} \sum x_{i1}^2 \sum x_{i1} x_{i2} \dots \sum x_{i1} x_{ij} \dots \sum x_{i1} x_{ip} \\ \sum x_{i2} \sum x_{i2} x_{i1} \sum x_{i2}^2 \dots \sum x_{i2} x_{ij} \dots \sum x_{i2} x_{ip} \\ \dots \\ \sum x_{ij} \sum x_{ij} x_{i1} \sum x_{ij} x_{i2} \dots \sum x_{ij}^2 \dots \sum x_{ij} x_{ip} \\ \dots \\ \sum x_{ip} \sum x_{ip} x_{i1} \sum x_{ip} x_{i2} \dots \sum x_{ip} x_{ij} \dots \sum x_{ip}^2 \end{pmatrix} \quad (33)$$

$$Y = \begin{pmatrix} y_1 \\ y_2 \\ \bullet \\ y_i \\ \bullet \\ y_n \end{pmatrix} \quad (34)$$

$$X^* Y = \begin{pmatrix} \sum y_i \\ \sum y_i x_{i1} \\ \sum y_i x_{i2} \\ \dots \\ \sum y_i x_{ij} \\ \dots \\ \sum y_i x_{ip} \end{pmatrix} \quad (35)$$

Pentru a rezolva sistemul de ecuații în formă de matrice (30), este necesar de a-l multiplica pe acesta de stânga și de dreapta cu matricea inversă matricei sistemului de ecuații normale:

$$(X^* X)^{-1} (X^* X)B = (X^* X)^{-1} (X^* Y) \quad (36)$$

$$(X^* X)^{-1} (X^* X) = E,$$

unde  $E$  este matricea cu diagonala principală formată din cifre 1.

Rezolvarea ecuației (30):

$$B = (X^* X)^{-1} (X^* Y) \quad (37)$$

Fiecare coeficient  $b_j$  va fi determinat cu formula:

$$b_j = \sum_{i=0}^n C_{ij} \sum_{i=1}^n y_i x_{ij} \quad (38),$$

unde  $C_{ij}$  sunt elementele matricei inverse –  $(X^* X)^{-1}$ .

Explicația cea mai clară a metodei pătratelor minime poate fi dată prin examinarea celui mai simplu caz, așa numita „regresia pereche”  $y = b_0 + b_1 x$  cu parametrizarea liniară. În acest caz formula (24) se scrie:

$$U = \sum_{i=1}^n [y_i - (b_0 + b_1 x_i)]^2 \quad (39)$$

Pentru a determina minimul funcției  $U$  se execută operațiunea:

$$\begin{cases} \frac{\partial U}{\partial b_0} = 0 \\ \frac{\partial U}{\partial b_1} = 0 \end{cases} \quad (40)$$

În amănunt:

$$\begin{cases} \frac{\partial U}{\partial b_0} = \sum_{i=0}^n [y_i - (b_0 + b_1 x_i)] = 0 \\ \frac{\partial U}{\partial b_1} = \sum_{i=1}^n [y_i - (b_0 + b_1 x_i)] x_i = 0 \end{cases} \quad (41)$$

sau:

$$\begin{cases} b_0 n + b_1 \sum x_i = \sum y_i \\ b_0 \sum x_i + b_1 \sum x_i^2 = \sum y_i x_i \end{cases} \quad (42)$$

Ecuatiile (42) se rezolvă cu ajutorul discriminanților:

$$b_0 = \frac{\Theta_1}{\Theta}, b_1 = \frac{\Theta_2}{\Theta}. \quad (43)$$

Discriminantul principal:

$$\begin{aligned} \Theta &= \begin{bmatrix} n \sum x & \sum x \sum x^2 \\ \sum x \sum x^2 & \sum x \sum x^3 \end{bmatrix} = n \sum x^2 - (\sum x)^2 \\ \Theta_1 &= \begin{bmatrix} \sum y \sum x & \sum xy \sum x^2 \\ \sum xy \sum x^2 & \sum xy \sum x^3 \end{bmatrix} = \sum y \sum x^2 - \sum xy \\ \Theta_2 &= \begin{bmatrix} n \sum y & \sum y \sum xy \\ \sum y \sum xy & \sum y \sum x^2 y \end{bmatrix} = n \sum xy - \sum x \sum y \end{aligned} \quad (44)$$

$$b_0 = \frac{\sum y \sum x^2 - \sum xy \sum x}{n \sum x^2 - (\sum x)^2} \quad (45)$$

$$b_1 = \frac{n \sum xy - \sum x \sum y}{n \sum x^2 - (\sum x)^2}$$

Pentru regresia inversă:

$$b_{0(xy)} = \frac{\sum x \sum y^2 - \sum xy \sum y}{n \sum y^2 - (\sum y)^2} \quad (46)$$

$$b_{1(xy)} = \frac{n \sum xy - \sum x \sum y}{n \sum y^2 - (\sum y)^2}$$

Discriminantul principal al doilea:

$$\Theta' = \begin{bmatrix} n \sum y & \sum y \sum y^2 \\ \sum y \sum y^2 & \sum y \sum y^3 \end{bmatrix} = n \sum y - (\sum y)^2. \quad (47)$$

Coeficientul de corelație la dependența pereche:

$$\begin{aligned} r &= \frac{\Theta_2}{\sqrt{\Theta \Theta'}} = (48) \\ &= \frac{n \sum xy - \sum x \sum y}{\sqrt{[n \sum x^2 - (\sum x)^2][n \sum y^2 - (\sum y)^2]}} \end{aligned}$$

Așa se aplică metoda pătratelor minime la analiza de regresie și corelație.

Ca exemplu este dată o problemă numerară mai complicată, care poate fi rezolvată numai cu ajutorul

calculatorului. Problema aceasta ține de construirea modelului matematic pentru calculul măsurii fluajului betonului în funcție de mai mulți factori cu evidența sensului fizic al fenomenului cercetat [1].

**Datele inițiale.** În calitate de funcție dependentă  $Y$  este analizată deformația specifică relativă (măsură curgerii lente) a betonului  $C_{(6,7)}$  în  $((\text{kg}/\text{cm}^2)^{-1}) \times 10^7$ . În calitate de factorii independenți  $x_j$  au fost analizați:

1.  $x_1$  – A – masa agregatelor betonului în kg/100.
2.  $x_2$  – C – masa cimentului în kg.
3.  $x_3$  – H – umiditatea aerului în %
4.  $x_4$  – S – factor-scară în cm.
5.  $x_5$  – W/C – raportul apă-ciment în %
6.  $x_6$  –  $\tau$  – vârsta betonului, zile
7.  $x_7$  – t – termenul de observații a deformațiilor betonului, zile.

Matricea de date inițiale  $X$  a avut 7 coloane și 260 de rânduri. Datele inițiale au fost acumulate din lucrările a 24 savanți.

Cu ajutorul programei REGRESS.FOR, elaborate la catedra elemente de construcții UTM, în calitate de **primul pas** al analizei, a fost construit modelul liniar:

$$y = -34,08856 - 2,79724x_1 + 0,23882x_2 - 1,20316x_3 - 0,61194x_4 + 3,11140x_5 - 0,23906x_6 + 0,01776x_7 \quad (49)$$

Caracteristicile statistice (suma pătratelor remanentă, coeficientul de corelație multiplă, criteriu Fișer) a acestui model sunt:

$$\begin{aligned} \text{SSR} &= 287001,0 ; R = 0,80248 ; F = 2,80875 \\ \text{valorile criteriului Student } t_j &: \\ t_1 &= -3,131 \quad t_2 = 6,062 ; t_3 = -9,631 ; \\ t_4 &= -1,719 ; t_5 = 16,027 ; t_6 = -5,460 ; \\ t_7 &= 6,394. \end{aligned}$$

Cu cât este mai mică SSR și cu cât mai mari sunt  $R$  și  $F$ , cu atât este mai bun modelul.

Criteriile  $t_j$  arată valoarea statistică a factorilor.

În **treapta a doua**, prin intermediul aceluiași program a fost construit modelul multiplicativ:

$$y = 0,01486 \times x_1^{-1,28955} \times x_2^{1,07747} \times x_3^{-1,08847} \times x_4^{-0,15233} \times x_5^{2,36261} \times x_6^{-0,19382} \times x_7^{0,32849} \quad (50)$$

Caracteristicile statistice (suma pătratelor remanentă, coeficientul de corelație multiplă, criteriu Fișer) a acestui model sunt:

$$\text{SSR} = 160200,5 ; R = 0,92427 ; F = 6,86239$$

Se știe că  $R$  variază de la 0 până la 1.

Acum despre sensul fizic. Teoria curgerii lente a betonului este o teorie veche. Mulți savanți au lucrat în domeniul acesta zeci de ani. Au fost construite mai multe teorii fenomenologice a fluajului betonului. Una din cele mai răspândite este teoria lui Mas-



lov – Arutiunian. În teoria aceasta măsura curgerii lente a betonului este descrisă în modul următor:

$$C(t, \tau) = \varphi(\tau) \times \Theta(t - \tau), \quad (51)$$

unde:

$$\varphi(\tau) = C_0 \times \Theta(\tau) \quad (52)$$

$$\Theta(\tau) = e^{-\gamma\tau} \quad (53) \quad \Theta(t - \tau) = e^{-\gamma(t-\tau)} \quad (54)$$

Autorul a propus că  $C_0$  să fie determinat cu formula:

$$C_0 = b_0 \prod_{j=1}^p x_j^{b_j} \quad (55),$$

unde:

$j \equiv 1, p$  este numărul de factori care influențează asupra  $C_0$ ,

$b_j$  – parametri,

$b_0$  – termenul liber al ecuației de regresie.

Cu aceste date se poate trece la **a treia treaptă**, la construcția modelului **multiplativ-expotent** al măsurii curgerii lente a betonului pe baza rezultatelor de observații (matricea datelor inițiale cu 7 coloane și 260 de rânduri). Pentru a transforma și retransforma membrii ecuației de regresie a fost pus în funcție programul PRE.FOR. Însă aici este o problemă: membrul  $[1 - e^{-\gamma(t-\tau)}]$  nu este comod să fie luat la o putere oarecare, de aceea procedura de calcul cu programele REGRES.FOR și PRE.FOR se repetă

de mai multe ori (metoda „încercări și greșeli”) până nu se obține semnul puterii la membru  $[1 - e^{-\gamma(t-\tau)}]$  egal cu 1. În așa fel a fost construit modelul matematic al măsurii curgerii lente a betonului cu evidența sensului fizic:

$$C(t, \tau) = 0,76996 * x_1^{-1,07862} * x_2^{0,75989} * x_3^{-1,22397} * x_4^{-0,21620} * x_5^{2,32528} * e^{-0,02 \times 0,2554\tau} * [1 - e^{-0,0072(t-\tau)}]^{1,0002} \quad (56).$$

Cum se vede din formula (56), semnul de putere la membrul  $[1 - e^{-\gamma(t-\tau)}]$  este egal cu  $1,0002 \approx 1$ . La acest model SSR

=134278,4. Cu alte cuvinte, SSR a fost micșorat cu 56% în comparație cu modelul liniar.

Așa dar, cu metoda pătratelor minime pot fi construite diferite modele matematice, cât se poate de complicate, liniare și neliniare, inclusiv cu evidența sensului fizic al fenomenului studiat.

### Literatura

1. Е. Львовский. Пассивный и активный эксперимент при изучении механических характеристик бетона. «Картеа Молдовенеаскэ», Кишинев, 1970.
2. Статистические методы построения эмпирических формул. «Высшая Школа», М., 1988.



Mihail Grecu. *Poarta privirii*. 1979. Pânză, tehnică mixtă