METODE DE PRELUCRARE A DATELOR EXPERIMENTALE

Membru corespondent al A.Ş.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.

Noţiunea „experimente pasive" necesită comentarii. Această noţiune a apărut în secolul trecut pentru a face diferenţă între experimente obişnuite clasice, *pasive*, în care în cursul experimentării se variază numai un singur factor ***x***, şi noul tip de experimente, *active,* în care în timpul experimentării se variază mai mulţi factori x_j conform unui plan special. Mulţi savanţi nu admit noţiunea „experimente pasive". Ei susţin că nu există niciun experiment pasiv. În fiecare experiment, chipurile, se implică activ cercetătorul. Poate că aşa şi este. Însă aceasta este terminologia ştiinţifică, acceptată în toată lumea.

Pe parcursul multor ani autorul a examinat o problema foarte actuală pentru ştiinţă. Este vorba de elaborarea modelelor matematice optimale pe baza datelor de observaţie pentru descrierea diferitelor fenomene. În urma acestor cercetări a fost obţinută o nouă metodă complexă cu multe trepte pentru construcţia modelelor, inclusiv cu evidenţa sensului fizic al fenomenelor studiate.

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

$$y = ax \tag{1}$$
$$y = a + bx \tag{2}$$
$$y = a + b/x \tag{3}$$
$$y = 1/(a + bx) \tag{4}$$
$$y = x/(a + bx) \tag{5}$$
$$y = ab^x \tag{6}$$
$$y = ae^{bx} \tag{7}$$
$$y = 1/(a + be^{-x} \tag{8}$$
$$y = ax^b \tag{9}$$
$$y = a + b\lg x \tag{10}$$
$$y = a/(b + x) \tag{11}$$
$$y = ax/(b + x) \tag{12}$$
$$y = ae^{b/x} \tag{13}$$
$$y = a + bx^n \tag{14}$$

Aşa formule pot fi multe, dar totuşi numărul lor este în cel mai simplu caz, care se studiază, limitat. De observat că toate formulele acestea pot fi liniarizate şi, ca rezultat, coeficienţii a, b pot fi determinaţi pe baza datelor de observaţii cu metoda pătratelor minime. În cazul, în care numărul factorilor $\boldsymbol{x_j}$ este mai mare decât 1 (j>1), numărul tipurilor de modele creşte la infinit [2].

Sistemul de ecuaţii, aşa numite normale, poate fi liniar, liniarizabil şi neliniar. În primele două cazuri pentru obţinerea modelelor matematice pe baza datelor experimentale poate fi utilizată metoda pătratelor minime. Ultimul caz e mai complicat. Aici, pentru a determina coeficienţii modelului matematic, este necesar de a rezolva un sistem de ecuaţii algebrice neliniare, ceea ce, după cum se ştie, nu e deloc simplu. Metoda pătratelor minime este un caz particular a metodei de verosimilitate maximală, când datele de observaţie sunt distribuite normal (aşa numita distribuţia lui Hauss). Dacă funcţia y (fenomenul examinat, variabila dependentă) este o valoare continuă (nu discretă), atunci probabilitatea faptului că aceasta funcţie în populaţia statistică generală va avea o oarecare valoare este:

$$P(y, \beta)dy, \tag{15}$$

unde β este un parametru necunoscut, care va fi determinat în cursul rezolvării problemei.

Formând un eşantion finit cu volumul *n* din populaţia statistică generală, obţinem valorile:

$$y_1, y_2, \ldots, y_n \tag{16}$$

Procesul de sustragere a membrilor eşantionului trebuie să fie aşa, ca probabilitatea obţinerii noii valori să fie independentă de valorile obţinute precedent. Probabilitatea de a obţine valorile (15) este egală:

$$dP = Ldy_1 dy_2 \ldots dy_n, \tag{17}$$

unde:

$$\begin{aligned} &L(y_1, y_2, \ldots, y_n; \beta) = \\ &= P\{y_1, \beta\} P\{y_2, \beta\} .. P\{y_n, \beta\}. \end{aligned} \tag{18}$$

Funcția L poartă numele de funcție de verosimilitate. Estimarea în eşantion a parametrului β 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:

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

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

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

Pentru un eşantion aleator cu volumul n se poate de scris:

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

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 şi metoda se numeşte metoda pătratelor minime.

Din aceasta se obține ecuația generală a metodei pătratelor minime:

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

unde $\eta\left(\vec{\beta}, \vec{x}\right)$ 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:

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

în care $1/\sigma_{y_i}^2 = W_i$ - coloniță-vector de greutăți specifice,

i – numărul încercărilor,

j – numărul factorilor în regresie multiplă.

Aşa 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 aşa numita dependență-pereche şi formula (23) se transformă şi se simplifică:

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

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

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

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

Prin rezolvarea acestor ecuații se determină coeficienții b_0 şi 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 $\boldsymbol{x_i}$ este atribuită numai o singură valoare a funcției ***y.*** Dispersia ***y*** nu există. Formula (26) se scrie:

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

Situația a doua. Fiecărei valori fixate a factorului $\boldsymbol{x_i}$ se atribuie ***k*** valori a variabilei dependente ***y,*** persistă distribuția normală, media teoretică η_i şi dispersia σ_{yi}^2 . Pentru a rezolva această problemă se foloseşte sistemul de ecuații (26). Persistă membrul $W = \frac{1}{\sigma_{yi}^2}.$

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 $\boldsymbol{k_x}$ valori ai factorului ***x*** cu distribuția normală, media teoretică ξ_i şi dispersia σ_{xi}^2. şi $\boldsymbol{k_y}$ valori pentru variabila dependentă ***y*** cu distribuția normală, media teoretică η_i şi dispersia σ_{yi}^2 . Distribuția ***x*** nu este corelată cu distribuția ***y.*** Pentru estimarea coeficienților β_0 şi β_j în acest caz metoda pătratelor minime nu este eficientă şi se aplică altă metodă, anume metoda ***confluentă.*** Metoda aceasta, fiind destul de complicată, se aplică de fapt numai în fizica nucleară şi aici nu se examinează.

Aşa dar, poate fi scrisă ecuaţia metodei pătratelor minime în cazul, când parametrizarea este liniară:

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

apoi se execută operaţiunea:

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

şi se obţine sistemul de ecuaţii a metodei pătratelor minime în cazul, când persistă mai mulţi factori $\boldsymbol{x_j}$ şi parametrizarea este liniară. În forma de matrice acest sistem se scrie:

$$\left(X^*X\right)B=X^*Y \qquad (30)$$

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

X – matricea tuturor valorilor factorilor examinaţi:

$$X=\begin{Vmatrix} x_{10}x_{11}x_{t2}\ldots x_{1j}\ldots x_{1p}\\ x_{20}x_{21}x_{22}\ldots x_{2j}\ldots x_{2p}\\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\\ x_{i0}x_{i1}x_{i2}\ldots x_{ij}\ldots x_{ip}\\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\\ x_{n0}x_{n1}x_{n2}\ldots x_{nj}\ldots x_{np} \end{Vmatrix} \qquad (32)$$

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

X^* - matricea sistemului de ecuaţii normale:

$$X^*=\begin{Vmatrix} n\sum x_{i1}\sum x_{i2}\ldots\sum x_{ij}\ldots\sum x_{ip}\\ \sum x_{i1}\sum x_{i1}^2\sum x_{i1}x_{i2}\ldots\sum x_{i1}x_{ij}\ldots\sum x_{i1}x_{ip}\\ \sum x_{i2}\sum x_{i2}x_{i1}\sum x_{i2}^2\ldots\sum x_{i2}x_{ij}\ldots\sum x_{i2}x_{ip}\\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\\ \sum x_{ij}\sum x_{ij}x_{i1}\sum x_{ij}x_{i2}\ldots\sum x_{ij}^2\ldots\sum x_{ij}x_{ip}\\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\\ \sum x_{ip}\sum x_{ip}x_{ii}\sum x_{ip}x_{i2}\ldots\sum x_{ip}x_{ij}\ldots\sum x_{ip}^2 \end{Vmatrix} \qquad (33)$$

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

$$X^*Y=\begin{Vmatrix} \sum y_i\\ \sum y_ix_{i1}\\ \sum y_ix_{i2}\\ \ldots\ldots\ldots\\ \sum y_ix_{ij}\\ \ldots\ldots\ldots\\ \sum y_ix_{ip} \end{Vmatrix} \qquad (35)$$

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

$$\begin{aligned}&\left(X^*X\right)^{-1}\left(X^*X\right)B=\left(X^*X\right)^{-1}\left(X^*Y\right)\\ &\left(X^*X\right)^{-1}\left(X^*X\right)=E,\end{aligned} \qquad (36)$$

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

Rezolvarea ecuaţiei (30):

$$B=\left(X^*X\right)^{-1}\left(X^*Y\right) \qquad (37)$$

Fiecare coeficient $\boldsymbol{b_j}$ va fi determinat cu formula:

$$b_j=\sum_{i=0}^{n}C_{ij}\sum_{i=1}^{n}y_ix_{ij} \qquad (38),$$

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

Explicaţia cea mai clară a metodei pătratelor minime poate fi dată prin examinarea celui mai simplu caz, aşa numita „regresia pereche” $y=b_0+b_1x$ cu parametrizarea liniară. În acest caz formula (24) se scrie:

$$U=\sum_{i=1}^{n}\left[y_i-\left(b_0+b_1x_i\right)\right]^2 \qquad (39)$$

Pentru a determina minimul funcţiei U se execută operaţiunea:

$$\left[\begin{aligned}\frac{\partial U}{\partial b_0}=0\\ \frac{\partial U}{\partial b_1}=0\end{aligned}\right. \qquad (40)$$

În amănunt:

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

sau:

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

Ecuațiile (42) se rezolvă cu ajutorul discriminanților:

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

Discriminantul principal:

$$\Theta = \begin{bmatrix} n \sum x \\ \sum x \sum x^2 \end{bmatrix} = n \sum x^2 - (\sum x)^2$$

$$\Theta_1 = \begin{bmatrix} \sum y \sum x \\ \sum xy \sum x^2 \end{bmatrix} = \sum y \sum x^2 - \sum xy \qquad (44)$$

$$\Theta_2 = \begin{bmatrix} n \sum y \\ \sum y \sum xy \end{bmatrix} = n \sum xy - \sum x \sum y$$

$$\begin{aligned} b_0 &= \frac{\sum y \sum x^2 - \sum xy \sum x}{n \sum x^2 - (\sum x)^2} \\ b_1 &= \frac{n \sum xy - \sum x \sum y}{n \sum x^2 - (\sum x)^2} \end{aligned} \qquad (45)$$

Pentru regresia inversă:

$$\begin{aligned} b_{0(xy)} &= \frac{\sum x \sum y^2 - \sum xy \sum y}{n \sum y^2 - (\sum y)^2} \\ b_{1(xy)} &= \frac{n \sum xy - \sum x \sum y}{n \sum y^2 - (\sum y)^2}. \end{aligned} \qquad (46)$$

Discriminantul principal al doilea:

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

Coeficientul de corelație la dependența pereche:

$$r = \frac{\Theta_2}{\sqrt{\Theta \Theta'}} = (48)$$

$$= \frac{n \sum xy - \sum x \sum y}{\sqrt{\left[n \sum x^2 - \left(\sum x \right)^2 \right] \left[n \sum y^2 - \left(\sum y \right)^2 \right]}}.$$

Așa se aplică metoda pătratelor minime la analiza de regresie şi 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. In calitate de funcție dependentă ***Y*** este analizată deformația specifică relativă (măsura curgerii lente) a betonului $C_{(t,\tau)}$ în $((\text{kg/cm}^2)^{-1})$x x 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 – τ – vârsta betonului, zile
7. x_7 – t – termenul de observații a deformațiilor betonului, zile.

Matricea de date inițiale **X** a avut 7 colonițe şi 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:

$$\begin{aligned} y &= -34{,}08856 - 2{,}79724x_1 + 0{,}23882x_2 - \\ &- 1{,}20316x_3 - 0{,}61194x_4 + 3{,}11140x_5 - \\ &- 0{,}23906x_6 + 0{,}01776x_7 \ (49). \end{aligned}$$

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

SSR=287001,0 ;R=0,80248; F=2,80875

valorile criteriului Stiudent $\mathbf{t_j}$:

t_1= -3,131 t_2=6,062; t_3= -9,631;

t_4= - 1,719 ; t_5 = 16,027 ; t_6= -5,460 ;

t_7 =6,394.

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

Criteriile $\mathbf{t_j}$ arată valoarea statistică a factorilor.

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

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

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

SSR=160200,5 ;R=0,92427; F=6,86239

Se ştie 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 Maslov – 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:

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

$$\Theta(\tau) = e^{-\gamma_1 \tau} \ (53) \quad \Theta(t-\tau) = e^{-\gamma(t-\tau)} \quad (54)$$

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:

$j \equiv \overline{1, p}$ 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 ***multiplicativ-expotent*** al măsurii curgerii lente a betonului pe baza rezultatelor de observații (matricea datelor inițiale cu 7 colonițe 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 $\left[1 - e^{-\gamma(t-\tau)}\right]$ 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 $\left[1 - e^{-\gamma(t-\tau)}\right]$ egal cu 1. În așa fel a fost construit modelul matematic al măsurii curgerii lente a betonului cu evidența sensului fizic:

$$C(t,\tau) = 0{,}76996 * x_1^{-1{,}07862} * x_2^{075989} * x_3^{-1{,}22397} * x_4^{-0{,}21620} * x_5^{2{,}32528} * e^{-0{,}02 \times 0{,}2554\tau} * \left[1 - e^{-0{,}0072(t-\tau)}\right]^{1{,}0002} \quad (56).$$

Cum se vede din formula (56), semnul de putere la membrul $\left[1 - e^{-\gamma(t-\tau)}\right]$ este egal cu 1,0002 ≈ 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***