APLICAREA METODEI CELOR MAI MICI PĂTRATE LA STUDIEREA CORELAȚIEI DINTRE FACTORII CLIMATERICI ÎN REPUBLICA MOLDOVA

Alina ȚURCANU, dr.
Universitatea Tehnică a Moldovei

Rezumat. Lucrarea prezintă modele matematice liniare și neliniare ce estimează evoluția proceselor sau fenomenelor pe baza unor parametri care definesc procesele şi fenomenele în vederea realizării de calcule şi aproximări ale datelor experimentale. Dispunând de o serie de date privind factorii climaterici, se prezintă analiza rezultatelor dintr-o perioadă de 126 ani, abordând metoda celor mai mici pătrate.
Cuvinte cheie: metoda celor mai mici pătrate, ecuația de regresie.

APPLYING THE METHOD OF LEAST SQUARES TO THE STUDY OF CORRELATION BETWEEN CLIMATE FACTORS IN THE REPUBLIC OF MOLDOVA

Abstract. The paper presents linear and nonlinear mathematical models that estimate the evolution of processes or phenomena based on parameters that define processes and phenomena in order to achieve calculations and approximations of experimental data. Featuring a series of climatic factors data, it is presented a 126-year analysis of the results, approaching the method of least squares.
Keywors: the method of least squares, the regression equation.

1. Introducere

Metoda celor mai mici pătrate (MCMMP) este folosita pentru a rezolva cu aproximare sisteme liniare și neliniare în care numărul de ecuații este mai mare decât numărul de necunoscute. MCMMP este folosită des în calcule statistice, în special in analiza de regresie.

MCMMP poate fi interpretată ca metodă de potrivire a datelor. Cea mai bună potrivire în sensul celor mai mici pătrate este acel model pentru care suma pătratelor valorilor reziduale este minimă, o valoare reziduala fiind diferența dintre o valoare bazată pe observație si o valoare dată de un model. MCMMP corespunde criteriului de risc maxim dacă erorile experimentale au o repartiție normală şi, totodată, poate fi interpretată ca metodă de estimare a momentelor.

Metoda celor mai mici pătrate îşi are originile pe tărâmul astronomiei și geodeziei, în încercarea oamenilor de ştiinţă şi a matematicienilor de a oferi soluţii de navigaţie pe oceane în timpul erei marilor descoperiri geografice. Descrierea precisă a comportamentului corpurilor cereşti a fost cheia ce a deschis calea navigaţiei pe oceane, unde marinarii nu mai aveau posibilitatea de a se ghida după poziţia uscatului. MCMMP reprezintă punctul culminant al unor cercetări ce au avut loc în secolul XVIII.

Metoda a fost descrisă pentru prima data de Carl Friedrich Gauss în jurul anului 1794. Matematicianul a pus bazele metodei celor mai mici pătrate în 1795, la vârsta de 18 ani. O primă demonstrație a puterii metodei lui Gauss a apărut când a fost folosită la prezicerea poziției viitoare a nou-descoperitului asteroid Ceres. Pe 1 ianuarie 1801, astronomul Giuseppe Piazzi a descoperit asteroidul Ceres și a reuşit sa-i urmărească traiectoria timp de 40 de zile, înainte de a-l pierde în strălucirea soarelui. Bazându-se pe aceste date, s-a dorit aflarea poziției lui Ceres după ce va apărea din spatele soarelui, fără a rezolva complicatele ecuații neliniare ale lui Kepler privind mişcarea planetelor. Singurele predicții care i-au permis astronomului maghiar Franz Xaver von Zach să determine cu succes poziția lui Ceres au fost cele realizate de Gauss, folosind analiza MCMMP.

Gauss a publicat metoda abia în 1809, în volumul doi al operei sale pe tema mecanicii cereşti, *„Theoria Motus Corporum Coelestium in sectionibus conicis solem ambientium"*. In 1829, Gauss a putut să afirme că apropierea dintre metoda celor mai mici pătrate şi analiza de regresie este optimă în sensul că, într-un model liniar în care erorile sunt necorelate, au media zero şi dispersii egale, cele mai bune estimări liniare nedeplasate ale coeficienților sunt estimările bazate pe MCMMP. Rezultatul este cunoscut drept Teorema Gauss-Markov.

În continuare, ne vom referi la situația regresiei liniare (relația dintre cele două variabile poate fi descrisă printr-o dreaptă în cadrul norului de puncte), parabolice și cubice. Regresia se leagă foarte mult de conceptul de corelație. Dacă am avea o corelație perfectă, estimarea ar fi extrem de precisă.

Republica Moldova este vulnerabilă la un şir de riscuri naturale cu impact mare asupra economiei şi societății. Acestea includ: eroziunile şi alunecările de teren, vânturile şi ploile puternice, secetele îndelungate, inundațiile devastatoare şi multe altele. Fiind o țară agrară, Moldova este afectată pe tot parcursul anului de diferite fenomene climatice de risc, care diminuează adesea puternic producția agricolă. De exemplu, uraganele puternice, seceta excesivă şi inundațiile vaste din vara anului 1994 au provocat numeroase jertfe omeneşti (47 de persoane) şi pagube materiale economiei naționale, estimate oficial la peste două miliarde lei moldoveneşti.

Scopul lucrării a fost de a prezenta cantitatea anuală a precipitațiilor în funcție de temperatura medie anuală a aerului. Pentru realizarea lui s-au folosit date statistice de la Serviciul Hidrometeorologic de Stat din Moldova.

2. Aproximarea prin metoda celor mai mici pătrate

Considerăm funcția reală $f: [a, b] \to \mathbb{R}$, pentru care sunt cunoscute valorile $y_i = f(x_i)$ în $(n+1)$ puncte distincte x_i, $i = \overline{o, n}$, din intervalul $[a, b]$, adică perechile de valori

$$(x_o, y_o); (x_1\ y_1); (x_2\ y_2); \ldots; (x_n\ y_n). \qquad (1)$$

În cazul general, punctele pot fi oarecare, dar ele sunt, de regulă, echidistante, cu pasul de discretizare h:

$$x_o = a,\ x_n = b,\ x_{i+1} - x_i = h = \frac{b-a}{n},\ \ i = \overline{o, n-1}. \qquad (2)$$

Se cere să se determine polinomul P_m, $grad\ P_m = m < n$, de forma

$$P_m(x) = c_0 + c_1 x + c_2 x^2 + \cdots + c_m x^m, \quad \forall x \in [a, b], \qquad (3)$$

care să aproximeze funcția f, astfel încît să fie minimizată suma pătratelor diferențelor dintre valorile aproximate și cele exacte în cele $(n+1)$ puncte. Altfel spus, trebuie rezolvată următoarea problemă de optimizare:

$$\hat{P}_m = \left\{P_m | \min_{c_0 \ldots c_n} J,\ \ J = \sum_{i=0}^{n} [P_m\ (x_i) \text{ -} y_i]^2\right\}. \qquad (4)$$

Metoda de calcul rezultată se numește **metoda celor mai mici pătrate (MCMMP)** și se utilizează atunci când fie perechile (1) nu sunt cunoscute cu exactitate, fie n este foarte mare.

Aproximarea funcției f cunoscute sub forma setului de valori (1) printr-un polinom de forma (3) prin MCMMP este numită în general și **regresie polinomială,** cu particularizările larg utilizate **regresie liniară** ($m = 1$), **regresie parabolică** ($m = 2$) și **regresie cubică** ($m = 3$). Aproximarea prin MCMMP poate fi aplicată însă și altor funcții de aproximare g, diferite de cele polinomiale.

2.1. ***Aproximarea polinomială liniară (m=1) prin MCMMP***

Algoritmul MCMMP ([1,2]) se bazează pe condiția că suma pătratelor diferențelor Δy_i sa fie minima, unde

$$\Delta y_i = y_i - f(c_0, c_1, x_i), \qquad (5)$$

$$S = \sum_i (\Delta y_i)^2 = min. \qquad (6)$$

Indicele fiecărei sume ia valori întregi în intervalul [1, n], n = numărul de valori x_i, respectiv y_i. Această metodă va fi aplicată după testarea nivelului erorilor și eliminarea erorilor grosolane.

Dependența funcțională se caută sub forma $y = c_0 + c_1 x$.

Pentru calculul lui c_0 și c_1 avem următorul sistem de ecuații:

$$\begin{cases} nc_0 + c_1 \sum X = \sum Y \\ c_0 \sum X + c_1 \sum X^2 = \sum XY, \end{cases} \qquad (7)$$

unde: n este numărul de cazuri cercetate; y este rezultatul estimat; c_0 este interceptul (locul pe ordonata unde dreapta de regresie se intersectează cu OY, valoarea lui y pentru $x = 0$); c_1 este panta de regresie (ne arată cu cât se modifică y atunci când x crește (scade) cu o unitate; x este variabilă criteriu (cunoscută).

Calcularea coeficienților de regresie c_0, respectiv c_1, conduce la realizarea primului pas din procesul regresiei.

Prin intermediul regresiei se pot face predicții ale unei variabile, în funcție de valoarea alteia. Predicția este procesul de estimare a valorii unei variabile cunoscând valoarea unei alte variabile.

2.2. ***Aproximarea polinomială parabolică (m=2) prin MCMMP***

Funcția polinomială de aproximare parabolică, numită și regresie parabolică, se caută sub forma $y = c_0 + c_1 x + c_2 x^2$.

Funcția J care trebuie minimizată, privită ca *funcție de variabilele* c_0, c_1, *și* c_2:

$$J = \sum_{i=0}^{n} (c_0 + c_1 x + c_2 x^2 - y_i)^2. \qquad (8)$$

Pentru minimizarea funcției convexe J este suficient să fie anulate derivatele sale parțiale, astfel obținând următorul sistem liniar de ecuații:

$$\begin{cases} (n+1)c_0 + c_1 \sum X + c_2 \sum X^2 = \sum Y \\ c_0 \sum X + c_1 \sum X^2 + c_3 \sum X^3 = \sum XY \\ c_0 \sum X^2 + c_1 \sum X^3 + c_3 \sum X^4 = \sum X^2 Y. \end{cases} \qquad (9)$$

2.3. ***Aproximarea polinomială cubică (m=3) prin MCMMP***

Funcția polinomială de aproximare cubică, numită și regresie cubică, se caută sub forma $y = c_0 + c_1 x + c_2 x^2 + c_3 x^3$. Funcția J care trebuie minimizată, privită ca *funcție de variabilele* c_0, c_1, c_2 *și* c_3:

$$J = \sum_{i=0}^{n} (c_0 + c_1 x + c_2 x^2 + c_3 x^3 - y_i)^2. \qquad (10)$$

Ca și în cazul aproximării polinomiale parabolice, pentru determinarea coeficienților c_0, c_1, c_2 și c_3 este necesar de următorul sistem liniar de ecuații:

$$\begin{cases} (n+1)c_0 + c_1 \sum X + c_2 \sum X^2 + c_3 \sum X^3 = = \sum Y \\ c_0 \sum X + c_1 \sum X^2 + c_2 \sum X^3 + c_3 \sum X^4 = \sum XY \\ c_0 \sum X^2 + c_1 \sum X^3 + c_2 \sum X^4 + c_3 \sum X^5 = \sum X^2 Y \\ c_0 \sum X^3 + c_1 \sum X^4 + c_2 \sum X^5 + c_3 \sum X^6 = \sum X^3 Y. \end{cases} \qquad (12)$$

3. Metoda celor mai mici pătrate la studierea corelației dintre temperatura medie anuală a aerului şi cantitatea anuală de precipitații

În acest paragraf, vom studia corelația dintre cantitatea anuală de precipitații şi anul calendaristic. Includem în tabelul de mai jos datele luate de la Serviciul Hidrometeorologic de Stat, precum şi variabilele ajutătoare necesare studiului ([3,4]):
X = Temperatura medie anuală a aerului; Y = cantitatea anuală de precipitații, mm.

n	Anul	X	Y		n	Anul	X	Y		n	Anul	X	Y
1	1891	9	430		43	1933	7,2	460		85	1975	9,8	500
2	1892	9,1	400		44	1934	9,9	520		86	1976	8,3	600
3	1893	10	440		45	1935	9,2	520		87	1977	9,5	470
4	1894	8,5	460		46	1936	10,3	520		88	1978	8,7	550
5	1895	9,8	450		47	1937	10	430		89	1979	9,7	680
6	1896	9,4	300		48	1938	10,2	330		90	1980	8,3	710
7	1897	9,8	560		49	1939	10,1	435		91	1981	9,7	560
8	1898	9,8	380		50	1940	9,5	520		92	1982	9,8	400
9	1899	9,7	520		51	1941	9,5	520		93	1983	10,5	565
10	1900	10,2	510		52	1942	9,5	520		94	1984	9,2	660
11	1901	9,7	500		53	1943	9,5	520		95	1985	8	600
12	1902	8,6	380		54	1944	9,2	350		96	1986	9,5	410
13	1903	9,9	510		55	1945	8,6	440		97	1987	8,2	600
14	1904	9	515		56	1946	10	530		98	1988	9	650
15	1905	9,7	520		57	1947	9,2	720		99	1989	10,9	500
16	1906	9,9	610		58	1948	9,3	620		100	1990	11,3	380
17	1907	8,4	520		59	1949	9,8	530		101	1991	9,5	680
18	1908	8,9	420		60	1950	10,2	400		102	1992	10,1	430
19	1909	9,2	525		61	1951	10,8	350		103	1993	9,5	520
20	1910	10,2	600		62	1952	10	600		104	1994	11,3	430
21	1911	9,4	530		63	1953	9,1	400		105	1995	10	700
22	1912	8,2	910		64	1954	8,8	560		106	1996	9,2	710
23	1913	9,5	440		65	1955	9,4	710		107	1997	9,5	610
24	1914	8,3	900		66	1956	8,4	500		108	1998	10,2	660
25	1915	9,5	520		67	1957	10,1	420		109	1999	11	500
26	1916	9,9	500		68	1958	10	590		110	2000	11,2	450
27	1917	10	515		69	1959	9,5	510		111	2001	10,4	610
28	1918	9,5	520		70	1960	10,6	520		112	2002	10,8	600
29	1919	8,9	520		71	1961	10,5	470		113	2003	9,8	470
30	1920	9,5	523		72	1962	10,1	560		114	2004	10,3	600
31	1921	9,3	430		73	1963	9,2	550		115	2005	10,5	638
32	1922	8,8	720		74	1964	9,3	510		116	2006	10,2	564
33	1923	10,2	530		75	1965	9	550		117	2007	12,1	480
34	1924	9	370		76	1966	10,9	770		118	2008	11,3	466
35	1925	10,5	440		77	1967	10	500		119	2009	11,4	446
36	1926	9,5	620		78	1968	10	520		120	2010	10,6	734
37	1927	9,7	520		79	1969	8,7	515		121	2011	10,5	428
38	1928	9	490		80	1970	10,1	680		122	2012	11,2	522
39	1929	8	400		81	1971	10	600		123	2013	11,1	531
40	1930	10,7	520		82	1972	9,8	610		124	2014	10,9	604
41	1931	8,8	520		83	1973	9,5	400		125	2015	12	431
42	1932	9	780		84	1974	10,9	550		126	2016	11,2	644
										Suma		**1226,1**	**66891**

n **=126**

$\sum X =$ **1226,1**

$\sum Y =$ **66891**

$\sum X^2 =$ **12024,57**

$\sum X^3 =$ **118837,13**

$\sum X^4 =$ **1183397,67**

$\sum X^5 =$ **11873169,96**

$\sum X^6 =$ **120011837,19**

$\sum XY =$ **649270**

$\sum X^2 Y =$ **6351479,99**

$\sum X^3 Y =$ **62612444,32**

3.1 Metoda celor mai mici pătrate polinomială liniară (m=1)

Dependența funcțională se caută sub forma $y = c_0 + c_1 x$. Sistemul (7) se va scrie:

$$\begin{cases} 126\, c_0 + 1226{,}1 c_1 = 66891 \\ 1226{,}1 c_0 + 12024{,}6 c_1 = 649270.3, \end{cases}$$

$$c_0 = 701{,}95, \text{ iar } c_1 = -17{,}58.$$

Ecuația de regresie obținută este:

$$y = 701{,}95 - 7{,}58\, x. \qquad (13)$$

Avem următoarele predicții: pentru temperatura medie anuală a aerului de 11 grade, vom avea cantitatea anuală de precipitații de 508,57 mm.

3. 2 Metoda celor mai mici pătrate polinomială parabolică (m=2)

Funcția polinomială de aproximare parabolică, numită și regresie parabolică, se caută sub forma $y = c_0 + c_1 x + c_2 x^2$. Sistemul (9) se va scrie:

$$\begin{cases} 126 c_0 + 1226{,}10\, c_1 + 12024{,}57\, c_2 = 66891{,}00 \\ 1226{,}10 c_0 + 12024{,}57\, c_1 + 118837{,}13\, c_2 = 649270{,}30 \\ 12024{,}57 c_0 + 118837{,}13\, c_1 + 1183397{,}67\, c_2 = 6351479{,}99. \end{cases}$$

Soluția sistemului: $c_0 = 687{,}35$, $c_1 = -14{,}57$, iar $c_2 = -0{,}15$. Regresia parabolică:

$$y = 687{,}35 - 14{,}57x - 0{,}15x^2. \qquad (14)$$

Respectiv, putem face următoarele predicții: pentru temperatura medie anuală a aerului de 11 grade, vom avea cantitatea anuală de precipitații de 483,38 mm.

3.3. *Metoda celor mai mici pătrate polinomială cubică (m=3)*

Funcția polinomială de aproximare cubică (regresie cubică) are forma:

$$y = c_0 + c_1 x + c_2 x^2 + c_3 x^3.$$

Sistemul (12) se va scrie:

$$\begin{cases} 126 c_0 + 1226{,}10\, c_1 + 12024{,}57\, c_2 + 118837{,}13 c_3 = 66891{,}00 \\ 1226{,}10 c_0 + 12024{,}57\, c_1 + 118837{,}13\, c_2 + 1183397{,}67 c_3 = 649270{,}30 \\ 12024{,}57 c_0 + 118837{,}13\, c_1 + 1183397{,}67\, c_2 + 11873169{,}96 c_3 = 6351479{,}99 \\ 118837{,}13 c_0 + 1183397{,}67\, c_1 + 11873169{,}96\, c_2 + 120011837{,}2 c_3 = 62612444{,}32. \end{cases}$$

Soluția sistemului este $c_0 = 2450{,}21$, $c_1 = -566{,}86$, iar $c_2 = 57{,}03$, iar $c_3 = -1{,}96$.

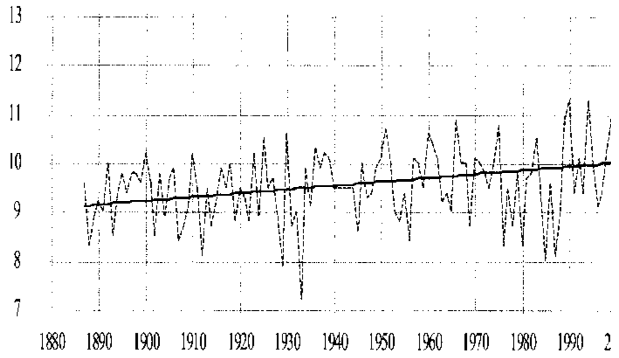
Regresia cubică este:

$$y = 2450{,}21 - 566{,}86x + 57{,}03x^2 - 1{,}96x^3. \qquad (15)$$

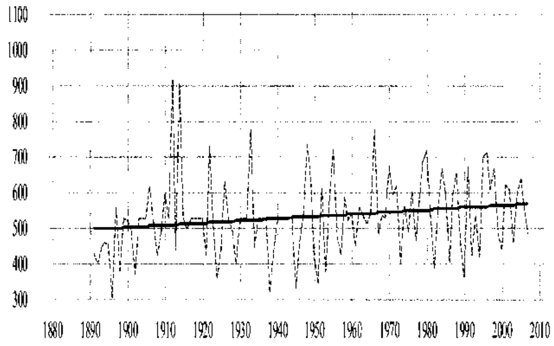
Respectiv, putem face următoarele predicții: pentru temperatura medie anuală a aerului de 11 grade, vom avea cantitatea anuală de precipitații de 510,5 mm.

Ca și în cazul aproximării polinomiale parabolice, pentru determinarea coeficienților c_0, c_1, c_2 și c_3 este necesar de următorul sistem liniar de ecuații:

Anexa 1. Temperatura medie anuală a aerului în oraşul Chişinău.



Anexa 2. Cantitatea anuală a precipitaţiilor în oraşul Chişinău.



6. Concluzii

Unul din principalele capitole ale statisticii are în vedere posibilitatea de a face predicţii. Deşi nu se găsesc relaţii perfecte în lumea reală, prin intermediul regresiei se pot face prognoze ale unei variabile, în funcţie de valoarea alteia.

Evident, rezultatele din lucrarea de faţă nu au un caracter determinist, deoarece temperatura medie anuală a aerului şi cantitatea anuală a precipitaţiilor mai depind şi de mulţi alţi factori, care nu pot fi luaţi în consideraţie, de exemplu, de poluarea accelerată a atmosferei, de managementul apelor, tehnologiile de utilizare a terenurilor etc.

Încălzirea globală a climei este considerată nu numai cel mai mare risc meteo-climatic, dar şi cel mai mare risc de mediu, ale cărei consecinţe negative se răsfrâng asupra tuturor geosferelor Terrei.

Numărul mare al investigaţiilor din ultimele decenii, atât în sfera fizică cât şi în cea biologică, şi interacţiunea lor cu schimbările climaterice la nivel regional şi naţional au dat posibilitate de a efectua o evaluare mai largă şi mai fermă a interacţiunilor dintre încălzirea observată şi consecinţele ei.

Principala concluzie este că asupra multor ecosisteme naturale şi artificiale influenţează schimbarea regională a climei, îndeosebi creşterea temperaturii şi intensificarea fenomenelor naturale de risc.

Schimbarea temperaturii şi cantităţii de precipitaţii va duce la modificarea perioadelor de vegetaţie, regimului higrologic al râurilor, la eroziunea solului, inundaţii, secete şi ploi torenţiale extrem de puternice.

Anume făcând anumite predicţii, putem să ne dăm seama de gravitatea situaţiei în care ne putem afla într-un moment, pentru a avea posibilitate de a preveni unele fapte şi întâmplări, care sunt catastrofale uneori, dacă nu sunt luate măsurile necesare la momentul potrivit.

Bibliografie

1. Piscunov N. S. Calculul diferenţial şi integral. Vol. 1. Chişinău: Ed. Lumina, 1992.
2. Ciumac P. ş.a. Teoria probabilităţilor şi elemente de statistică matematică. Chişinău: Editura „Tehnica” UTM, 2003.
3. Научно прикладной справочник по климату СССР (Молдавская СССР), серия 3. Многолетние данные. Часть 1-6. Выпуск II. Ленинград: Гидрометеоздат, 1990.
4. http://www.meteo.md.