B A D A N I A   O P E R A C Y J N E   I   D E C Y Z J E

Nr 2 2004

Agata KORCZAK*

APROKSYMACJA SKOŃCZONEGO UKŁADU PUNKTÓW PŁASZCZYZNY FIGURĄ BĘDĄCĄ SUMĄ PUNKTU I PROSTEJ

Przedstawiono nową metodę aproksymacji skończonego układu punktów płaszczyzny, zawierającego punkty „odstające" od układu. Dotychczas proponowane metody eliminowały te punkty, a pozostały układ aproksymowały prostą. W opisanej metodzie nie odrzuca się żadnego punktu. Jako krzywą aproksymującą przyjęto figurę będącą sumą punktu i prostej i wyznaczono jej parametry. Otrzymana w ten sposób prosta – nazwana odporną prostą regresji – jest uniezależniona od „odstających" punktów, które opisuje wyznaczony punkt. Metodę zilustrowano przykładami.

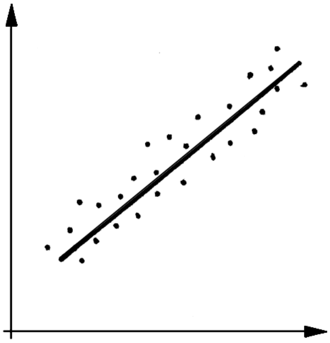
Słowa kluczowe: *aproksymacja, krzywa regresji, funkcja liniowa, funkcja kwadratowa, odporność krzywej regresji, iteracja*

1. Wstęp

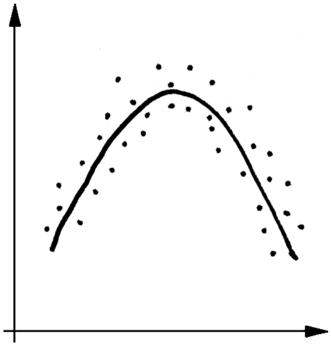
Ważnym zagadnieniem w matematyce i praktyce jest, mając dane pewne punkty – np. na płaszczyźnie – znaleźć przepis funkcji, krzywej, która je opisuje. Dobieramy krzywą, która najlepiej opisuje położenie punktów, najlepiej je aproksymuje. Punkty te często obrazują wyniki doświadczeń, rezultaty badań. Dobór odpowiedniej krzywej jest zatem istotny dla możliwości przewidywania kolejnych wyników. Istnieje kilka metod znajdywania krzywej najlepiej aproksymującej [1, s. 128–130]. Najczęściej używaną jest metoda najmniejszych kwadratów, wymaga ona bowiem jedynie znajomości podstaw rachunku różniczkowego. Chcąc wyznaczyć krzywą najlepiej aproksymującą dany układ punktów, czyli krzywą regresji, należy przede wszystkim określić rodzinę krzywych, wśród której szukamy krzywej regresji. Podstawowym kryterium, którym kierujemy się przy wyborze odpowiedniej rodziny krzywych jest

* Politechnika Wrocławska, filia w Wałbrzychu, ul. Armii Krajowej 78, 59-220 Wałbrzych, e-mail: leswoj@poczta.onet.pl

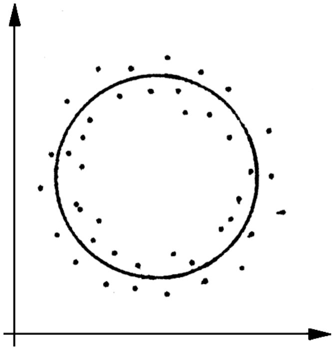
tzw. metoda wzrokowa – choć naturalnie może mieć ona zastosowanie jedynie do punktów w przestrzeni co najwyżej trójwymiarowej [2, s. 18].



Rys. 1

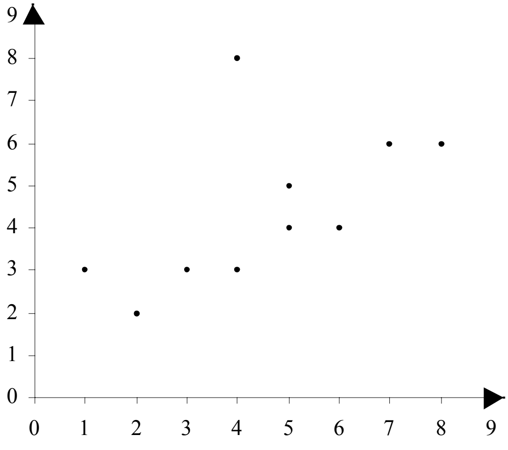


Rys. 2



Rys. 3

Przyglądając się na przykład rozmieszczeniu punktów na rysunku 1, nie mamy wątpliwości, że funkcja liniowa jest krzywą regresji opisującą dany układ. Chcąc znaleźć jej przepis, szukalibyśmy współczynników m, n w równaniu $y = mx + n$. Przykład z rysunku 2 sugeruje, że krzywą regresji jest tu funkcja kwadratowa (a nie, jak wcześniej, liniowa), należałoby więc znaleźć współczynniki a, b, c funkcji $y = ax^2 + bx + c$. Rysunek 3 obrazuje rozkład punktów wokół okręgu, więc dochodzimy do wniosku, że układ ten jest opisany przez zależność niejawną $(x - a)^2 + (y - b)^2 = r^2$, należałoby zatem znaleźć współczynniki a, b, r.



Rys. 4

Innym pytaniem, które zadajemy sobie rozważając krzywe regresji, jest pytanie o odporność, czyli jaki wpływ na przebieg krzywej mają punkty, które w jakiś sposób

„odstają" od układu (rys. 4). Krzywa regresji jest bowiem krzywą, która tylko najlepiej opisuje układ punktów, a nie krzywą, do której należą punkty. Problemem tym zajmowali się między innymi w swoich monografiach Huber [3] i Hampel [4]. W dotychczasowych metodach próbowano „oczyścić" układ z „odstających" punktów eliminując je bądź traktując jako anomalie, mało istotnie wpływające na kształt prostej regresji [4, s. 344–374].

W przedstawionej metodzie zauważamy „odstawanie" niektórych punktów układu i dlatego aproksymujemy figurą będącą sumą punktu i prostej. Punkt „ściągnie" do siebie wszystkie punkty odstające od układu, a prosta opisze pozostały układ. Krzywą regresji (y względem x) będziemy szukać zatem nie wśród rodziny prostych, ale wśród rodziny krzywych płaskich będących sumą mnogościową punktu i prostej. Krzywe te opisuje zależność

$$\sqrt{(x-a)^2+(y-b)^2}\,(y-mx-n)=0, \tag{1}$$

gdzie a, b, m, n są parametrami rzeczywistymi.

2. Obliczenie parametrów

Rozważymy na płaszczyźnie skończony zbiór punktów A, zbiór przynajmniej dwuelementowy, o punktach nie leżących na jednej prostej prostopadłej do osi odciętych:

$$A=\{(x_1,y_1),(x_2,y_2),\ldots(x_n,\ y_n)\}=\begin{bmatrix} x_1 & x_2 & \cdots & x_n \\ y_1 & y_2 & \cdots & y_n \end{bmatrix}.$$

W celu znalezienia krzywej (1) posłużymy się metodą Antoniewicza, która umożliwia szukanie regresji opisanych zależnościami niejawnymi [6, s. 22].

Zadanie sprowadza się do znalezienia minimum wyrażenia

$$\Delta(a,b,m,n,A)=\sum_A\left(\sqrt{(x-a)^2+(y-b)^2}\right)^2(y-mx-n)^2, \tag{2}$$

gdzie: $\Delta(a, b, m, n, A)$ oznacza miarę niespełnienia równania (1) przez elementy zbioru A, a, b, m, n – współczynniki szukanej krzywej.

Dla lepszej czytelności zapisów wprowadzamy oznaczenie: $\sum\limits_A = \sum$. Równanie (2) można wtedy napisać w postaci

$$\Delta(a,b,m,n,A)=\Sigma\,((x-a)^2+(y-b)^2)\,(y-mx-n)^2. \tag{3}$$

Warunkiem koniecznym istnienia minimum powyższego wyrażenia jest zerowanie się pochodnych cząstkowych pierwszego rzędu względem każdej ze zmiennych [7, s. 368]. Mamy:

$\Delta'_a = \Sigma 2(-1)(x-a)(y-mx-n)^2 = 2\Sigma(a-x)(y^2+m^2x^2+n^2-2mxy-2ny+2mnx)$
$= 2\Sigma(ay^2+am^2x^2+an^2-2amxy-2any+2amnx-xy^2-m^2x^3-n^2x+2mx^2y$
$+ 2nxy-2mnx^2)$
$= 2(a[y^2]+am^2[x^2]+an^2[1]-2am[xy]-2an[y]+2amn[x]-[xy^2]-m^2[x^3]$
$- n^2[x]+2m[x^2y]+2n[xy]-2mn[x^2]),$

$\Delta'_b = 2\Sigma(b-y)(y^2+m^2x^2+n^2-2mxy-2ny+2mnx)$
$= 2\Sigma(by^2+bm^2x^2+bn^2-2bmxy-2bny+2bmnx-y^3-m^2x^2y-n^2y+2mxy^2+2ny^2$
$- 2mnxy)$
$= 2(b[y^2]+bm^2[x^2]+bn^2[1]-2bm[xy]-2bn[y]+2bmn[x]-[y^3]-m^2[x^2y]$
$- n^2[y]+2m[xy^2]+2n[y^2]-2mn[xy]),$

$\Delta'_m = \Sigma 2((x-a)^2+(y-b)^2)(y-mx-n)(-x) = 2\Sigma(x^2-2ax+a^2+y^2-2by+b^2)$
$\times (mx^2+nx-xy)$
$= 2\Sigma(mx^4+nx^3-x^3y-2amx^3-2anx^2+2ax^2y+a^2mx^2+a^2nx-a^2xy+mx^2y^2+nxy^2$
$- xy^3-2\mathrm{b}mx^2y-2bnxy+2bxy^2+mb^2x^2+nb^2x-b^2xy)$
$= 2(m([x^4]+[x^2y^2])+m[x^2](a^2+b^2)+n[x](a^2+b^2)-[xy](a^2+b^2)+n([x^3]$
$+ [xy^2])-2am[x^3]-2an[x^2]+2a[x^2y]-2bm[x^2y]-2bn[xy]+2b[xy^2]-[x^3y]-[xy^3]),$

$\Delta'_n = \Sigma 2((x-a)^2+(y-b)^2)(y-mx-n)(-1) = 2\Sigma(x^2-2ax+a^2+y^2-2by+b^2)$
$\times (mx+n-y) = 2\Sigma(mx^3+nx^2-x^2y-2amx^2-2anx+2axy+a^2mx+a^2n$
$- a^2y+mxy^2+ny^2-y^3-2bmxy-2bny+2by^2+mb^2x+nb^2-b^2y)$
$= 2(m([x^3]+[xy^2])+m[x](a^2+b^2)+n[1](a^2+b^2)-[y](a^2+b^2)+n([x^2]$
$+ [y^2])-2am[x^2]-2an[x]+2a[xy]-2bm[xy]-2bn[y]+2\mathrm{b}[y^2]-[x^2y]-[y^3]).$

Powyższe pochodne zerują się wtedy i tylko wtedy, gdy:

$$a[y^2]+am^2[x^2]+an^2[1]-2am[xy]-2an[y]+2amn[x] \\ -m^2[x^3]-n^2[x]+2m[x^2y]+2n[xy]-2mn[x^2] = [xy^2], \tag{4}$$

$$b[y^2]+bm^2[x^2]+bn^2[1]-2bm[xy]-2bn[y]+2bmn[x] \\ -m^2[x^2y]-n^2[y]+2m[xy^2]+2n[y^2]-2mn[xy] = [y^3], \tag{5}$$

$$m([x^4]+[x^2y^2])+m[x^2](a^2+b^2)+n[x](a^2+b^2)-[xy](\mathrm{a}^2+b^2)+n([x^3]+[xy^2]) \\ -2am[x^3]-2an[x^2]+2a[x^2y]-2bm[x^2y]-2bn[xy]+2b[xy^2] = [x^3y]+[xy^3], \tag{6}$$

$$m([x^3]+[xy^2])+m[x](a^2+b^2)+n[1](a^2+b^2)-[y](a^2+b^2)+n([x^2]+[y^2]) \\ -2am[x^2]-2an[x]+2a[xy]-2bm[xy]-2bn[y]+2\mathrm{b}[y^2] = [x^2y]+[y^3]. \tag{7}$$

Otrzymaliśmy cztery równania z czterema zmiennymi, każde stopnia trzeciego. Łatwo zauważyć, że równanie (4) jest liniowe ze względu na zmienną *a,* nie występuje w nim natomiast zmienna *b*. Analogicznie równanie (5) jest liniowe ze względu na *b*, brak w nim natomiast zmiennej *a*. Z równań (4) i (5) można zatem wyliczyć zmienne *a*, *b*:

$$a = \frac{n^2[x] + m^2[x^3] - 2m[x^2y] - 2n[xy] + 2mn[x^2] + [xy^2]}{[y^2] + m^2[x^2] + n^2[1] - 2m[xy] - 2n[y] + 2mn[x]},$$

$$b = \frac{n^2[y] + m^2[x^2y] - 2m[xy^2] - 2n[y^2] + 2mn[xy] + [y^3]}{[y^2] + m^2[x^2] + n^2[1] - 2m[xy] - 2n[y] + 2mn[x]}.$$

Równania (6), (7), które są liniowe ze względu na *m*, *n*, zapiszemy w postaci:

$$m([x^2](a^2 + b^2) + [x^4] + [x^2y^2] - 2a[x^3] - 2b[x^2y]) + n([x](a^2 + b^2) + [x^3] + [xy^2] - 2a[x^2] - 2b[xy]) = [xy](a^2 + b^2) - 2a[x^2y] - 2b[xy^2] + [x^3y] + [xy^3], \tag{8}$$

$$m([x](a^2 + b^2) + [x^3] + [xy^2] - 2a[x^2] - 2b[xy]) + n([1](a^2 + b^2) + [x^2] + [y^2] - 2a[x] - 2b[y]) = [y](a^2 + b^2) - 2a[xy] - 2b[y^2] + [x^2y] + [y^3]. \tag{9}$$

Dla łatwiejszych rachunków oznaczamy:

$K = [x^2](a^2 + b^2) + [x^4] + [x^2y^2] - 2a[x^3] - 2b[x^2y]$,
$L = [x](a^2 + b^2) + [x^3] + [xy^2] - 2a[x^2] - 2b[xy]$,
$M = [xy](a^2 + b^2) - 2a[x^2y] - 2b[xy^2] + [x^3y] + [xy^3]$,
$N = [x](a^2 + b^2) + [x^3] + [xy^2] - 2a[x^2] - 2b[xy]$,
$O = [1](a^2 + b^2) + [x^2] + [y^2] - 2a[x] - 2b[y]$,
$P = [y](a^2 + b^2) - 2a[xy] - 2b[y^2] + [x^2y] + [y^3]$.

Z równań (8) i (9) wyliczamy *m*, *n*:

$$m = \frac{MO - LP}{KO - LN}, \quad n = \frac{KP - MN}{KO - LN},$$

gdzie $KO - LN \neq 0$.

Układ równań (4)–(7) przybiera zatem postać:

$$a = \frac{n^2[x] + m^2[x^3] - 2m[x^2y] - 2n[xy] + 2mn[x^2] + [xy^2]}{[y^2] + m^2[x^2] + n^2[1] - 2m[xy] - 2n[y] + 2mn[x]}, \tag{10}$$

$$b = \frac{n^2[y] + m^2[x^2y] - 2m[xy^2] - 2n[y^2] + 2mn[xy] + [y^3]}{[y^2] + m^2[x^2] + n^2[1] - 2m[xy] - 2n[y] + 2mn[x]}, \tag{11}$$

$$m = \frac{MO - LP}{KO - LN}, \quad n = \frac{KP - MN}{KO - LN}. \tag{12}$$

Algebraiczne, ścisłe rozwiązanie tego układu jest trudne. Dużo łatwiej rozwiązywać go metodą iteracyjną, odnosząc się do konkretnych przykładów. Wstawiamy „na chybił trafił” wartości a, b do równań (12) i wyliczamy m, n. Z kolei te wyniki wstawiamy do równań (10), (11) i otrzymujemy nowe wartości a, b. Powtarzamy te czynności tak długo, aż uznamy, że wartości a, b są dostatecznie dokładne. Znajdujemy w ten sposób w przybliżeniu punkt stacjonarny ($\boldsymbol{a}_0$, $\boldsymbol{b}_0$, $\boldsymbol{m}_0$, $\boldsymbol{n}_0$) wyrażenia (3). Jak już wcześniej zauważyliśmy, każde z równań (4), (5), (6), (7) jest liniowe ze względu na jedną zmienną, istnieje więc tylko jedno rozwiązanie układu. Należy upewnić się, czy w otrzymanym punkcie stacjonarnym wyrażenie (3) ma minimum. W tym celu obliczamy wszystkie pochodne cząstkowe drugiego rzędu funkcji $\Delta(a, b, m, n, A)$:

$$\begin{aligned}
&\Delta''_{aa} = 2([y^2] + m^2[x^2] + n^2[1] - 2m[xy] - 2n[y] + 2mn[x]),\\
&\Delta''_{ab} = \Delta''_{ba} = 0,\\
&\Delta''_{am} = \Delta''_{ma} = 2(2am[x^2] - 2a[xy] + 2an[x] - 2m[x^3] + 2[x^2y] - 2n[x^2]),\\
&\Delta''_{an} = \Delta''_{na} = 2(2an[1] - 2a[y] + 2am[x] - 2n[x] + 2[xy] - 2m[x^2]),\\
&\Delta''_{bb} = 2([y^2] + m^2[x^2] + n^2[1] - 2m[xy] - 2n[y] + 2mn[x]),\\
&\Delta''_{bm} = \Delta''_{mb} = 2(2bm[x^2] - 2b[xy] + 2bn[x] - 2m[x^2y] + 2[xy^2] - 2n[xy]),\\
&\Delta''_{bn} = \Delta''_{nb} = 2(bn[1] - 2b[y] + 2bm[x] - 2n[y] + 2[y^2] - 2m[xy]),\\
&\Delta''_{mm} = 2([x^4] + [x^2y^2] + [x^2](a^2 + b^2) - 2a[x^3] - 2b[x^2y]),\\
&\Delta''_{mn} = \Delta''_{nm} = 2([x](a^2 + b^2) + [x^3] + [xy^2] - 2a[x^2] - 2b[xy]),\\
&\Delta''_{nn} = 2([1](a^2 + b^2) + [x^2] + [y^2] - 2a[x] - 2b[y]).
\end{aligned}$$

Macierz drugiej pochodnej (13) oraz jej podwyznaczniki główne (13a,b,c,d) mają więc postać:

$$\Delta'' = \begin{bmatrix} \Delta''_{aa} & \Delta''_{ab} & \Delta''_{am} & \Delta''_{an} \\ \Delta''_{ba} & \Delta''_{bb} & \Delta''_{bm} & \Delta''_{bn} \\ \Delta''_{ma} & \Delta''_{mb} & \Delta''_{mm} & \Delta''_{mn} \\ \Delta''_{na} & \Delta''_{nb} & \Delta''_{nm} & \Delta''_{nn} \end{bmatrix}. \tag{13}$$

$$W_1 = \Delta''_{aa}, \tag{13a}$$

$$W_2 = \begin{vmatrix} \Delta''_{aa} & \Delta''_{ab} \\ \Delta''_{ba} & \Delta''_{bb} \end{vmatrix}, \tag{13b}$$

$$W_3 = \begin{bmatrix} \Delta''_{aa} & \Delta''_{ab} & \Delta''_{am} \\ \Delta''_{ba} & \Delta''_{bb} & \Delta''_{bm} \\ \Delta''_{ma} & \Delta''_{mb} & \Delta''_{mm} \end{bmatrix}, \tag{13c}$$

$$W_4 = \begin{bmatrix} \Delta''_{aa} & \Delta''_{ab} & \Delta''_{am} & \Delta''_{an} \\ \Delta''_{ba} & \Delta''_{bb} & \Delta''_{bm} & \Delta''_{bn} \\ \Delta''_{ma} & \Delta''_{mb} & \Delta''_{mm} & \Delta''_{mn} \\ \Delta''_{na} & \Delta''_{nb} & \Delta''_{nm} & \Delta''_{nn} \end{bmatrix}. \tag{13d}$$

Jak wiadomo z kryterium Syvestera, jeśli wszystkie podwyznaczniki główne (13a, b, c, d) w punkcie ($\boldsymbol{a}_0$, $\boldsymbol{b}_0$, $\boldsymbol{m}_0$, $\boldsymbol{n}_0$) są dodatnie, to wyrażenie (3) osiąga w punkcie stacjonarnym minimum [8, s. 373], a co za tym idzie układ A aproksymuje krzywa:

$$\sqrt{(x-\boldsymbol{a}_0)^2+(y-\boldsymbol{b}_0)^2}\cdot(y-\boldsymbol{m}_0 x-\boldsymbol{n}_0)=0 \,.$$

Prosta $y - \boldsymbol{m}_0 x - \boldsymbol{n}_0 = 0$ została nazwana **odporną prostą regresji** (y względem x) [10].

3. Aproksymacja wybranych układów punktów

Przykład 1

Odczytując dane z rysunku 4, otrzymamy

$$A = \begin{bmatrix} 1 & 2 & 3 & 4 & 4 & 5 & 5 & 6 & 7 & 8 \\ 3 & 2 & 3 & 3 & 8 & 4 & 5 & 4 & 6 & 6 \end{bmatrix}.$$

Patrząc na rozkład punktowy tego układu dochodzimy do wniosku, że krzywą, która najlepiej aproksymuje zbiór A, jest prosta: $y - m_1 x - n_1 = 0$, gdzie współczynniki m_1, n_1 można obliczyć metodą najmniejszych kwadratów [9, s.229–231]. Otrzymujemy:

$$m_1 = \frac{[xy][1]-[x][y]}{[x^2][1]-[x^2]}, \quad n_1 = \frac{[x^2][y]-[x][xy]}{[x^2][1]-[x^2]}, \tag{14}$$

gdzie:

$$[x^p y^r] = \sum_{i=1}^{10} x_i^p y_i^r ,$$

$p, r = 0, 1, 2, ...$

w szczególności:

$$[xy]=\sum_{i=1}^{10} x_i y_i,$$

$$[1]=\sum_{i=1}^{10} x^0 y^0 = 10.$$

W celu dokonania potrzebnych obliczeń stworzymy tabelę pomocniczą.

Tabela 1

Numer	Współrzędne punktów		x^2	xy
	x	y		
1	1	3	1	3
2	2	2	4	4
3	3	3	9	9
4	4	3	16	12
5	4	8	16	32
6	5	4	25	20
7	5	5	25	25
8	6	4	36	24
9	7	6	49	42
10	8	6	64	48
$[1] = 10$	$[x] = 45$	$[y] = 44$	$[x^2] = 245$	$[xy] = 219$

Ź r ó d ł o: opracowanie własne (dane umowne).

Po podstawieniu do wzorów (14) otrzymamy:

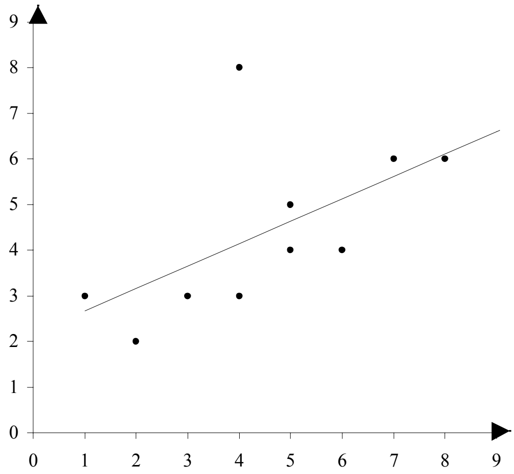
$$m_1 = 0{,}494118, \quad n_1 = 2{,}176471.$$

Prostą regresji y względem x określa zatem wzór

$$y - 0{,}4941x - 2{,}1765 = 0, \tag{15}$$

a obrazuje rysunek 5.

Łatwo można zauważyć, że punkt, nazwijmy go $P_1 = \begin{bmatrix} 4 \\ 8 \end{bmatrix}$, zdecydowanie odstaje od całego układu. Aproksymujemy zatem układ A nie prostą, ale sumą mnogościową prostej i punktu. W tym celu stworzymy tabelę pomocniczą z wartościami danymi występującymi w równaniach (10), (11), (12).



Rys. 5

Tabela 2

Numer	Punkty		x^2	xy	y^2	x^3	x^2y	xy^2	y^3	x^4	x^2y^2	x^3y	xy^3
	x	y											
1	1	3	1	3	9	1	3	9	27	1	9	3	27
2	2	2	4	4	4	8	8	8	8	16	16	16	16
3	3	3	9	9	9	27	27	27	27	81	81	81	81
4	4	3	16	12	9	64	48	36	27	256	144	192	108
5	4	8	16	32	64	64	128	256	512	256	1024	512	2048
6	5	4	25	20	16	125	100	80	64	625	400	500	320
7	5	5	25	25	25	125	125	125	125	625	625	625	625
8	6	4	36	24	16	216	144	96	64	1296	576	864	384
9	7	6	49	42	36	343	294	252	216	2401	1764	2058	1512
10	8	6	64	48	36	512	384	288	216	4096	2304	3072	1728
10	**45**	**44**	**245**	**219**	**224**	**1485**	**1261**	**1177**	**1286**	**9653**	**6943**	**7923**	**6849**

Ź r ó d ł o: opracowanie własne (dane umowne).

Tabela 3 obrazuje szukanie wartości zmiennych a, b, m, n. Jako początkowe wartości a, b wybieramy współrzędne punktu $P_1 = \begin{bmatrix} 4 \\ 8 \end{bmatrix}$. Punkt ten nie spełnia jednak naszych równań, lecz już po kilku iteracjach jesteśmy w stanie podać wartości a, b, m, n z dokładnością do czterech miejsc po przecinku.

Tabela 3

Dane zadawane		*m*	*n*	Dane wyliczone	
a	*b*			*a*	*b*
4	8	0,5093	1,5832	4,0451	7,3885
4,01	7,3	0,5122	1,5952	4,0376	7,3789
4,03	7,37	0,5109	1,5953	4,0400	7,3811
4,04	7,38	0,5105	1,5963	4,0408	7,3816
4,0407	7,381	0,5104	1,5963	4,0408	7,3816
4,0408	7,3815	0,5104	1,5963	4,0409	7,3816
4,0409	**7,3816**	0,5104	1,5963	**4,0409**	**7,3816**

Ź r ó d ł o: opracowanie własne (dane umowne).

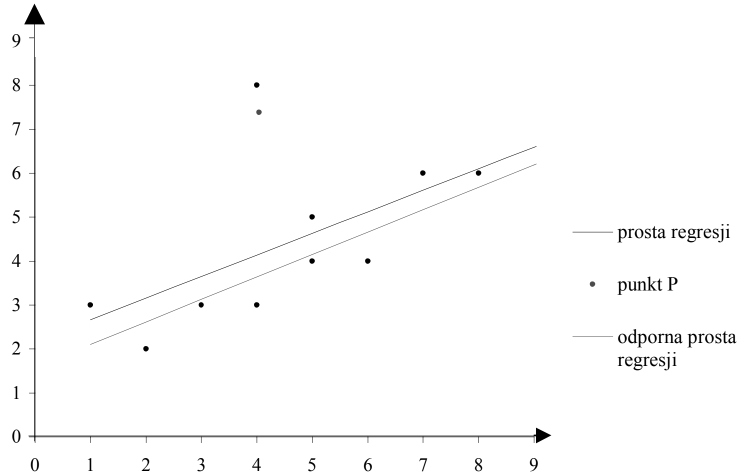
Znaleźliśmy zatem w przybliżeniu punkt stacjonarny ($\boldsymbol{a}_0$, $\boldsymbol{b}_0$, $\boldsymbol{m}_0$, $\boldsymbol{n}_0$) wyrażenia (3):

$$\boldsymbol{a}_0 = 4{,}0409, \quad \boldsymbol{b}_0 = 7{,}3816, \quad \boldsymbol{m}_0 = 0{,}5104, \quad \boldsymbol{n}_0 = 1{,}5963.$$

Wartości poszczególnych podwyznaczników (13a, b, c, d) wynoszą:
$W_1 = 45{,}2047$
$W_2 = 2043{,}4625$
$W_3 = 12928734{,}45$
$W_4 = 1064577954$

Oczywiście, wszystkie są dodatnie. Pokazaliśmy więc, że w otrzymanym punkcie stacjonarnym wyrażenie (3) ma minimum. Układ *A* można zatem aproksymować krzywą:

$$\sqrt{(x-4{,}0409)^2+(y-7{,}3816)^2}\,(y-0{,}5104x-1{,}5963)=0.$$



Rys. 6

Punkt $P_1 = \begin{bmatrix} 4{,}0409 \\ 7{,}3816 \end{bmatrix}$ zaspokaja w układzie wszystkie odstające punkty, a prosta $y - 0{,}5104x - 1{,}5963 = 0$ jest prostą regresji opisującą pozostałe punkty (rys. 6).

Prosta $y - 0{,}5104x - 1{,}5963 = 0$ jest **odporną prostą regresji** (y względem x), natomiast prosta $y - 0{,}4941x - 2{,}1765 = 0$ jest **zwyczajną prostą regresji** (y względem x) [10] (rys. 6).

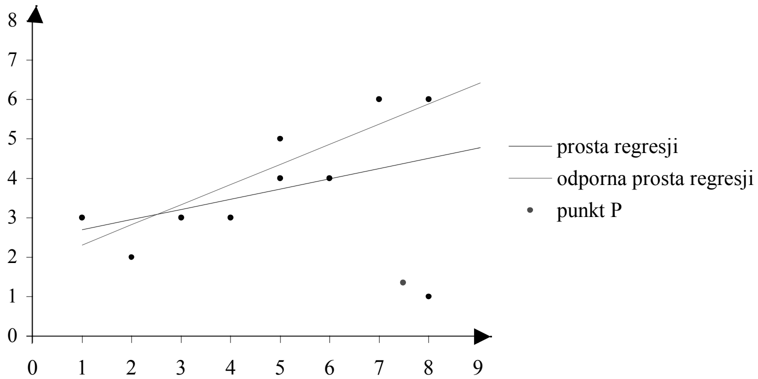
Przykład 2

$$B = \begin{bmatrix} 1 & 2 & 3 & 4 & 5 & 5 & 6 & 7 & 8 & 8 \\ 3 & 2 & 3 & 3 & 4 & 5 & 4 & 6 & 1 & 6 \end{bmatrix}.$$

Przy aproksymacji prostą $y - m_1x - n_1 = 0$ prosta regresji ma postać

$$y - 0{,}2590x - 2{,}4310 = 0.$$

Aproksymując punktem i prostą $\sqrt{(x-a)^2 + (y-b)^2}\ (y - mx - n) = 0$, otrzymujemy odporną prostą regresji $y - 0{,}5112x - 1{,}7979 = 0$ oraz punkt $P = \begin{bmatrix} 7{,}4816 \\ 1{,}3473 \end{bmatrix}$.



Rys. 7

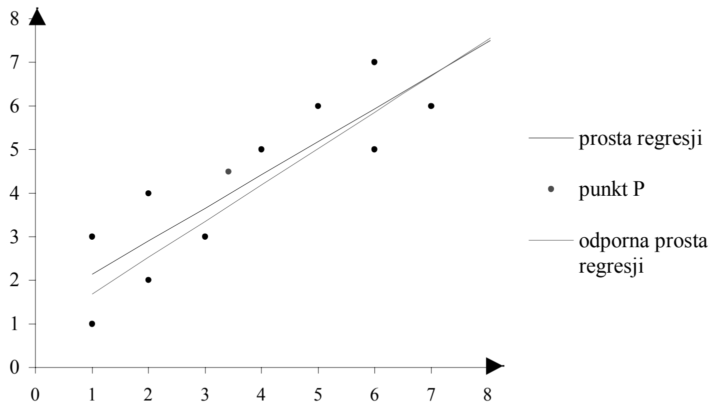
Przykład 3

$$C = \begin{bmatrix} 1 & 1 & 2 & 2 & 3 & 4 & 5 & 6 & 6 & 7 \\ 1 & 3 & 2 & 4 & 3 & 5 & 6 & 5 & 7 & 6 \end{bmatrix}.$$

Przy aproksymacji prostą $y - m_1x - n_1 = 0$ prosta regresji ma postać

$$y - 0{,}7619x - 1{,}3809 = 0.$$

Aproksymując punktem i prostą $\sqrt{(x-a)^2+(y-b)^2}$ (y – mx – n) = 0, otrzymujemy odporną prostą regresji $y - 0{,}8324x - 0{,}8593 = 0$ oraz punkt $P = \begin{bmatrix} 3{,}4217 \\ 4{,}4899 \end{bmatrix}$.



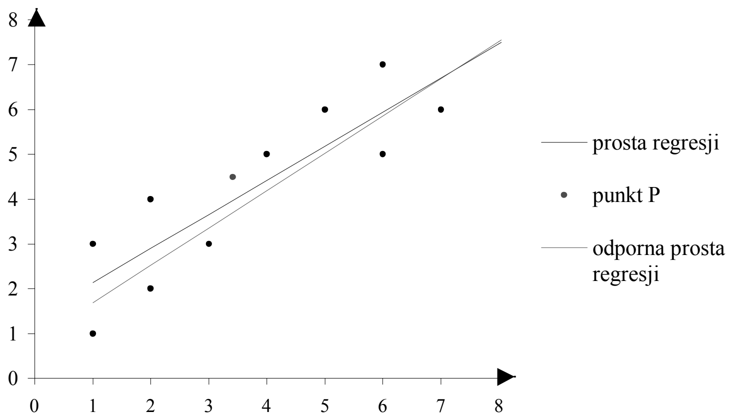
Rys. 8

Przykład 4

$$D = \begin{bmatrix} 1 & 2 & 3 & 3 & 3 & 5 & 6 & 8 & 8 & 9 \\ 9 & 3 & 2 & 3 & 5 & 5 & 7 & 7 & 9 & 3 \end{bmatrix}.$$

Przy aproksymacji prostą $y - m_1x - n_1 = 0$ prosta regresji ma postać

$$y - 0{,}1760x - 4{,}4553 = 0.$$



Rys. 9

Aproksymując punktem i prostą $\sqrt{(x-a)^2+(y-b)^2}$ $(y - mx - n) = 0$, otrzymujemy odporną prostą regresji $y + 0{,}2478x - 8{,}6733 = 0$ oraz punkt $P = \begin{bmatrix} 3{,}6987 \\ 3{,}2394 \end{bmatrix}$.

4. Wnioski

Już z obserwacji tych czterech prostych przykładów można stwierdzić, że punkty „zanieczyszczające" układ mocno wpływają na przebieg prostej regresji.

Układy punktów A i B różnią się tylko punktami odstającymi (rys. 6, 7). Odporne proste regresji różnią się niewiele, można przyjąć, iż są równoległe:

$y - \mathbf{0{,}5104}x - 1{,}5963 = 0$ odporna prosta regresji dla układu A,
$y - \mathbf{0{,}5112}x - 1{,}7979 = 0$ odporna prosta regresji dla układu B.

Zwyczajne proste regresji mają natomiast zupełnie inny przebieg:

$y - \mathbf{0{,}49}x - 2{,}18 = 0$ prosta regresji dla układu A,
$y - \mathbf{0{,}259}x - 2{,}431 = 0$ prosta regresji dla układu B.

W układzie C zauważamy, że nie ma punktu, który istotnie odstaje od pozostałych, stąd odporna prosta regresji niewiele różni się od zwyczajnej prostej regresji (rys. 8).

Patrząc zaś na odporną prostą regresji i zwyczajną prostą regresji układu D, stwierdzamy zupełnie odmienny ich przebieg:

$y - \mathbf{0{,}1760}x - 4{,}4553 = 0$,
$y + \mathbf{0{,}2478}x - 8{,}6733 = 0$.

Zadajemy sobie kolejne pytanie: skoro można aproksymować układ punktem i prostą, to czy nie byłoby korzystniej dla układu D aproksymować go prostą i dwoma punktami? Jest to kolejny temat, nad którym warto się zastanowić, wykracza on jednak poza ramy tej pracy.

Bibliografia

[1] ANTONIEWICZ R., *Metoda najmniejszych kwadratów dla zależności niejawnych i jej zastosowania w ekonomii*, Prace Naukowe Akademii Ekonomicznej we Wrocławiu Nr 445, Wrocław 1988.
[2] ANTONIEWICZ R., *O odpornej prostej regresji*, Zeszyty Naukowe Nr 320, Uniwersytet Szczeciński, Szczecin 2001.
[3] FICHTENCHOLZ G.M., *Rachunek różniczkowy i całkowy*, t. I, PWN, Warszawa 1985.

[4] HAMPEL F.R., RONCHETTI E.M., ROUSSEEUW P. J., STAHEL W.A., *Robus Statistics*, The Approach Based on Influence Functions, Wiley, New York 1981.
[5] HELLWIG Z., *Aproksymacja stochastyczna*, PWE, Warszawa 1965.
[6] HUBER P.J., *Robust Statistics*, Wiley, New York 1981.
[7] SOBCZYK M., *Statystyka*, PWN, Warszawa 2001.
[8] ZAJĄC K., *Zarys metod statystycznych*, PWE, Warszawa 1988.

Approximation of plane point finite system by a figure being the sum of point and straight line

Approximation of the plane point finite system is not only a theoretical problem. It exists also in engineering and economy. The points on the plane are the result of investigation and experience. So, it is highly probable to obtain results – points that are "not congruent" with the others. Sometimes it happens just accidentally, it may, however, be also the evidence of some anomalies.

In this article, we consider a finite plane points the system dispersed along the straight line, including the points "not congruent" with the system. While seeking the regression curve describing a given plane point system, we often ask ourselves to what extent the curve course is influenced by those "not congruent" points. The method proposed previously assumed that the curve course is not influenced substantially by those "not congruent" points and approximated the whole system by a straight line or eliminated those points and the remaining system was approximated by the straight line. In the method described below no point is rejected. As an approximating curve there was assumed a figure being the sum of point and straight lines, and afterwards its parameter were determined. The straight line obtained in this way – called the resistant regression line – does not depend on the "not congruent" points described by the point determined. The method is illustrated by examples calculating each time – for comparison – the residstant regression straight line and the regression straight line.

Keywords: *approximation, regression curve, linear function, quadratic function, regression curve resistance, iteration*