

Pewna procedura stosowania liniowej wielowymiarowej analizy regresji do słabo uwarunkowanych danych eksperymentalnych

Streszczenie: W artykule przedstawiono pewien nowy sposób obliczeń parametrów modelu regresji. Polega on na odpowiednim sprowadzeniu regresji wielowymiarowej do regresji dwuwymiarowej poprzez wprowadzenie tzw. zmiennych zastępczych. Dzięki temu można łatwo wykryć wadliwe wyrazy, które mogą występować w macierzy obserwacji, a także bez trudu zweryfikować liniowość modelu regresyjnego. Jest to istotne zwłaszcza w przypadku tzw. układów słabo uwarunkowanych. Sposób ten i jego walory przedstawiono na licznych elementarnych przykładach numerycznych.

Słowa kluczowe: analiza regresji, słabo uwarunkowane układy, zastosowania inżynierskie.

A certain method of multivariate regression analysis application to poorly conditioned experimental data

Summary: The article presents a new method of the regression model parameters calculation. The procedure consists in a suitable reduction of multivariate regression analysis to two-dimensional regression by means of creating so-called substitute variables. Due to that it is easy to detect defective elements which can exist in the observation matrix and linearity of regression model can be easily verified. This is especially essential in case of so-called poorly conditioned systems. The method and its advantages are presented in a lot of elementary numerical examples.

Keywords: regression analysis, poorly conditioned systems, engineering application.

1. Wprowadzenie

W badaniach eksperymentalnych dość często mamy do czynienia z sytuacją, gdy zależność pewnej mierzonej wielkości y od zbioru pewnej liczby p tzw. wielkości niezależnych x_1, x_2, \dots, x_p modelujemy za pomocą wielowymiarowej funkcji liniowej postaci:

$$y = b_0 + b_1x_1 + b_2x_2 + \dots + b_px_p \quad (1)$$

Dysponujemy przy tym pewnym zbiorem wartości zmierzonych, tworzących tzw. macierz obserwacji postaci:

$$[x_{1i}, x_{2i}, \dots, x_{pi}, y_i] \quad (2)$$

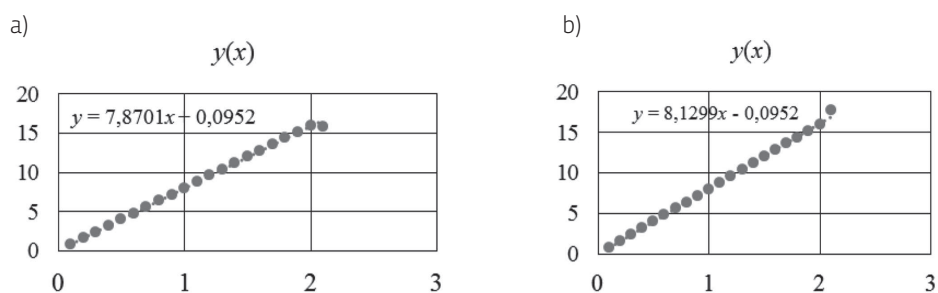
gdzie $i = 1, 2, \dots, n$ zaś n jest liczbą obserwacji. Zagadnienie to jest powszechnie znane pod nazwą wielowymiarowej analizy regresji. Istnieje bogata literatura na ten temat, która opisuje tę metodę głównie od strony teoretycznej, traktując związek (1) jako związek zmiennych stochastycznych [1–17]. Wartości optymalne nieznanymi parametrów stałych b_μ ($\mu = 0, 1, 2, \dots, p$), występujących w modelu (1) wyznacza się w powszechnie znany sposób wykorzystujący kryterium najmniejszej sumy kwadratów błędów [1, 16], zaś cała procedura obliczeń doczekała się bogatego oprogramowania i jest doskonale znana [12, 13, 17]. Jednakże praktyczne stosowanie tej metody, zwłaszcza przy dość dużej liczbie zmiennych niezależnych (znaczącej wartości liczby p), daje niejednokrotnie dość podejrzane wyniki. Dzieje się tak zwłaszcza wtedy, gdy wiemy, że określone parametry b_μ powinny przyjmować wartości dodatnie, zaś korzystając z obliczeń metody analizy regresyjnej otrzymujemy wartości ujemne.

Dla przykładu, wiemy, że masy, stałe sprężystości czy współczynniki tłumienia w dynamice układów mechanicznych nie mogą przyjmować wartości ujemnych.

Jak się wydaje, wymienić można trzy główne tego przyczyny:

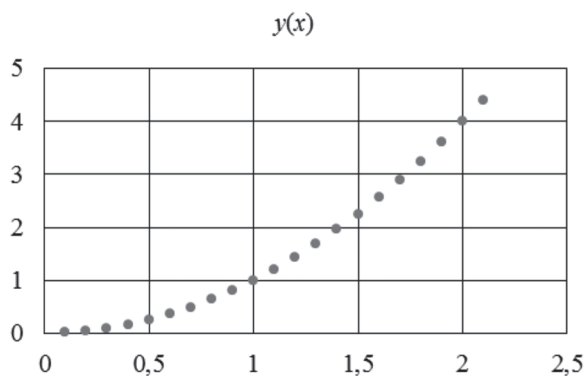
- pomiary niektórych wierszy macierzy obserwacji są obarczone zbyt dużym błędem (niektóre obserwacje są wadliwe),
- równania uzyskane przez podstawienie poszczególnych wierszy macierzy obserwacji (2) do modelu (1) są częściowo liniowo zależne (tzw. słabe uwarunkowanie – dla $n = p + 1$ wyznacznik główny układu jest bliski zeru),
- postać przyjętego modelu nie jest liniowa (jest inna niż równanie (1)).

Dla przykładu, wystarczy drobny błąd w jednej wartości zmiennej y funkcji $y = 8x$, aby uzyskać ujemną bądź dodatnią stałą b_0 (por. rys. 1).



Rys. 1. Przykład wpływu drobnego błędu w jednej wartości zmiennej y funkcji $y = 8x$ na wyniki regresji

Podobnie, jeśli funkcja $y(x)$ jest nieliniowa (np. $y = x^2$), uzyskuje się ujemną wartość stałej b_0 (por. rys. 2).



Rys. 2. Przykład wpływu nieliniowości zależności $y(x)$ na wyniki regresji

Jak widać, w przypadku dwuwymiarowym ($p = 1$) nie jest trudno zorientować się, który punkt (lub punkty) psują wyniki, a także czy przyjęty model liniowy może być stosowany. W tym celu wystarczy zwykła obserwacja wykresu zależności eksperymentalnej $y(x)$ (dla przykładu z rysunku 1 od razu wiadomo, że punktem wadliwym jest punkt 21, uzyskany dla maksymalnej wartości zmiennej x , zaś dla przykładu z rysunku 2 od razu wiadomo, że zależność $y(x)$ jest nieliniowa typu parabolicznego). Sytuacja komplikuje się dla przypadku wielowymiarowego. Już dla $p > 2$ nie sposób jest stworzyć wykresów o odpowiedniej liczbie zmiennych, a tym samym łatwo ocenić, które punkty są wadliwe lub z jakiego rodzaju nieliniowością mamy do czynienia.

Dla przykładu dla funkcji dwuwymiarowej $y(x_1, x_2)$ postaci:

$$y = 0,6 + 0,4x_1 + 0,3x_2 \quad (3)$$

dla $x_1 = 1, 2, \dots, 8$, przyjmując dodatkowo związek $x_2 = 3x_1$, otrzymujemy macierz obserwacji postaci:

$$A = \begin{bmatrix} 1 & 3 & 1,9 \\ 2 & 6 & 3,2 \\ 3 & 9 & 4,5 \\ 4 & 12 & 5,8 \\ 5 & 15 & 7,1 \\ 6 & 18 & 8,4 \\ 7 & 21 & 9,7 \\ 8 & 24 & 11 \end{bmatrix} \quad (4)$$

Niestety, dla tej macierzy obserwacji, stosując profesjonalne oprogramowanie, nie otrzyma się zadanych wartości $b_0 = 0,6$; $b_1 = 0,4$; $b_2 = 0,3$. Zmienne x_1, x_2 są bowiem liniowo zależne. Wprowadzenie drobnego błędu w przypadku chociażby jednej obserwacji (np. obserwacji dla $x_1 = 7$) powoduje, że wyniki są już możliwe do uzyskania, jednakże różnią się one znacznie od wartości występujących w funkcji (3) (por. rys. 3). Układ ten jest bowiem słabo uwarunkowany. W układzie takim drobna zmiana pojedynczej wartości istotnie wpływa na wyniki. Widać, że w rozpatrywanym przykładzie stałe b_1, b_2 mogą być zarówno dodatnie, jak i ujemne.

```
A:matrix([1,3,1.9],[2,6,3.2],[3,9,4.5],[4,12,5.8],[5,15,7.1],[6,18,8.4],[7,21,9.7],[8,24,11])$
fpprintprec:45
wyn:linear_regression(A);
expt: undefined: 0 to a negative exponent.
#0:
linear_regression(dat=matrix([1,3,1.9],[2,6,3.2],[3,9,4.5],[4,12,5.8],[5,15,7.1],[6,18,8.4],[7,21,9.7],[8,24,11]),select=[])
(C:\maxima-5.37.2\share\maxima\5.37.2_4_g0743c29_dirty\share\stats\stats.mac line 1277)
-- an error. To debug this try: debugmode(true);
C:matrix([1,3,1.9],[2,6,3.2],[3,9,4.5],[4,12,5.8],[5,15,7.1],[6,18,8.4],[7,21,5,9.01],[8,24,11])$
```

```
(%i7)
      LINEAR REGRESSION MODEL
      b_estimation= [0.6 , 5.44 , -1.38 ]
      b_statistics= [2.691 1010 , 2.734 1010 , -2.102 1010 ]
      b_p_values= [0.0 , 0.0 , 0.0 ]
      b_distribution= [ student_t , 5 ]
      v_estimation= 7.824 10-22
      v_conf_int= [3.049 10-22 , 4.706 10-21 ]
      v_distribution= [ chi2 , 5 ]
      adc=1.0
```

```
D:matrix([1,3,1.9],[2,6,3.2],[3,9,4.5],[4,12,5.8],[5,15,7.1],[6,18,8.4],[7,21,5,10.02],[8,24,11])$
```

```
(%i9)
      LINEAR REGRESSION MODEL
      b_estimation= [0.6 , -0.62 , 0.64 ]
      b_statistics= [1.606 1011 , -1.86 1010 , 5.818 1010 ]
      b_p_values= [0.0 , 0.0 , 0.0 ]
      b_distribution= [ student_t , 5 ]
      v_estimation= 2.197 10-23
      v_conf_int= [8.56 10-24 , 1.321 10-22 ]
      v_distribution= [ chi2 , 5 ]
      adc=1.0
```

Rys. 3. Przykład wyników analizy regresji dla układu słabo uwarunkowanego (wg programu wxMaxima)

przy czym $v = 2, 3, \dots, p$. Nowe zmienne Y, X są zmiennymi zastępczymi. Jak można za-
uważyć, liczba zmiennych niezależnych uległa redukcji z p do $p - 1$. Wyjściowy układ n
równań postaci (4) można więc zastąpić układem $n - 1$ równań postaci (7), w którym liczba
nieznanymi i zmiennych niezależnych jest o jeden mniejsza.

W ten sposób można dalej postępować sukcesywnie, aż do uzyskania dowolnie małej
liczby zmiennych niezależnych w równaniu regresji, a szczególności do uzyskania postaci
dwuwymiarowej:

$$Y = b_{p-1} + b_p X. \quad (9)$$

3. Przykłady numeryczne dla $p = 2$

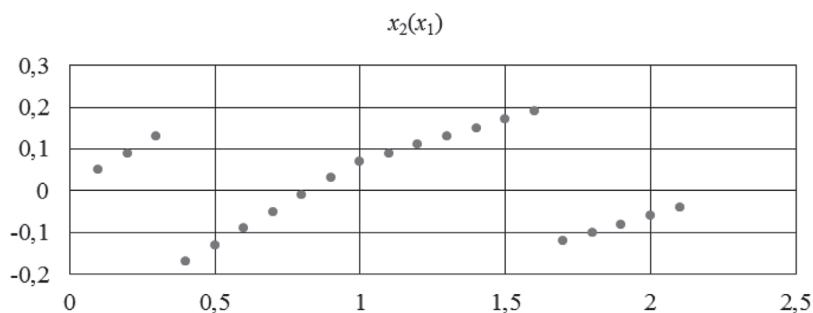
Przedstawiony powyżej sposób został sprawdzony na kilku przykładach układów do-
brze uwarunkowanych i słabo uwarunkowanych. W pierwszym przykładzie utworzono
macierz obserwacji \mathbf{B} za pomocą funkcji liniowej postaci:

$$y = 1,4 + 10x_1 + 13x_2. \quad (10)$$

Wartości zmiennej niezależnej x_1 zmieniano w przedziale od 0,1 do 2,1, równomiernie
z krokiem kwantowania 0,1. Przyjęto więc, że macierz \mathbf{B} ma 21 wierszy ($i = 1, 2, \dots, 21$) i trzy
kolumny: x_1, x_2, y . Wartości zmiennej x_2 zadano tak, aby uzyskać częściową niezależność
zmiennych x_1, x_2 (por. rys. 4). Dla tak przyjętych wartości zmiennych x_1, x_2 obliczono war-
tości wyrazów kolumny y dokładnie według funkcji (10). W ten sposób w przykładzie tym
powinno się uzyskać idealnie dokładne wartości stałych b_0, b_1, b_2 . Przyjęto więc następu-
jącą postać macierzy \mathbf{B} :

$$\mathbf{B} = ([0,1,0,05,3,05],[0,2,0,09,4,57],[0,3,0,13,6,09],[0,4,-0,17,3,19],[0,5,-0,13,4,71], (11) [0,6,-0,09,6,23], [0,7,-$$

 $0,05,7,75],[0,8,-0,01,9,27],[0,9,0,03,10,79],[1,0,0,7,12,31], [1,1,0,09,13,57],[1,2,0,11,14,83],[1,3,0,13,16,09],[1,4,0,15,17,3$
 $5],[1,5,0,17,18,61], [1,6,0,19,19,87],[1,7,-0,12,16,84],[1,8,-0,1,18,1],[1,9,-0,08,19,36],[2,-0,06,20,62],[2,1,-0,04,21,88])$

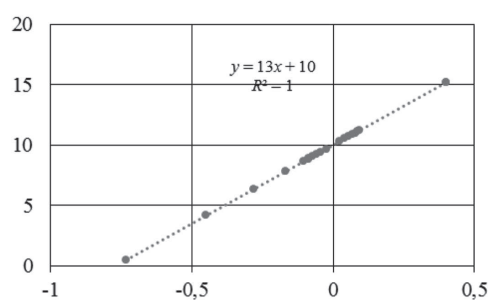


Rys. 4. Wykres zależności zmiennych x_2, x_1 w macierzy obserwacji \mathbf{B} postaci (11)

Stosując dla tej macierzy opisany powyżej sposób redukcji zmiennych, uzyskano ide-
alnie dokładne wyniki (por. rys. 5). Obliczenia wykonano, stosując typowe oprogramowa-
nie pakietu Excel 2013.

x_1	x_2	y	Δx_1	Δx_2	$X=\Delta x_2 / \Delta x_1$	$Y=\Delta y / \Delta x_1$	y	$z=13x_2 + 10x_1$	$y-z$
0,1	0,05	3,05	-	-	-	-	-	-	-
0,2	0,09	4,57	0,1	0,04	0,4	15,2	4,57	3,17	1,4
0,3	0,13	6,09	0,2	0,08	0,4	15,2	6,09	4,69	1,4
0,4	-0,17	3,19	0,3	-0,22	-0,733333	0,466667	3,19	1,79	1,4
0,5	-0,13	4,71	0,4	-0,18	-0,45	4,15	4,71	3,31	1,4
0,6	-0,09	6,23	0,5	-0,14	-0,28	6,36	6,23	4,83	1,4
0,7	-0,05	7,75	0,6	-0,1	-0,16667	7,833333	7,75	6,35	1,4
0,8	-0,01	9,27	0,7	-0,06	-0,08571	8,885714	9,27	7,87	1,4
0,9	0,03	10,79	0,8	-0,02	-0,025	9,675	10,79	9,39	1,4
1	0,07	12,31	0,9	0,02	0,022222	10,28889	12,31	10,91	1,4
1,1	0,09	13,57	1	0,04	0,04	10,52	13,57	12,17	1,4
1,2	0,11	14,83	1,1	0,06	0,054545	10,70909	14,83	13,43	1,4
1,3	0,13	16,09	1,2	0,08	0,066667	10,86667	16,09	14,69	1,4
1,4	0,15	17,35	1,3	0,1	0,076923	11	17,35	15,95	1,4
1,5	0,17	18,61	1,4	0,12	0,085714	11,11429	18,61	17,21	1,4
1,6	0,19	19,87	1,5	0,14	0,093333	11,21333	19,87	18,47	1,4
1,7	-0,12	16,84	1,6	-0,17	-0,10625	8,61875	16,84	15,44	1,4
1,8	-0,1	18,1	1,7	-0,15	-0,08824	8,852941	18,1	16,7	1,4
1,9	-0,08	19,36	1,8	-0,13	-0,07222	9,061111	19,36	17,96	1,4
2	-0,06	20,62	1,9	-0,11	-0,05789	9,247368	20,62	19,22	1,4
2,1	-0,04	21,88	2	-0,09	-0,045	9,415	21,88	20,48	1,4

$Y(X)$



WYNIKI:

- $b_0 = 1,4$
- $b_1 = 10$
- $b_2 = 13$

Rys. 5. Obliczenia wartości parametrów b_0, b_1, b_2 – sposobem redukcji liczby zmiennych dla macierzy B postaci (11)

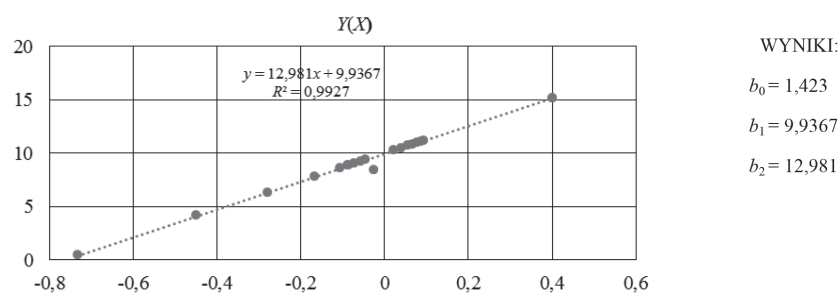
Identyczne wyniki uzyskano, stosując dla tej macierzy profesjonalne oprogramowanie statystyczne zawarte w pakiecie wxMaxima (por. rys. 6).

```

LINEAR REGRESSION MODEL
b_estimation= [1.4 , 10.0 , 13.0 ]
b_statistics= [1.94 1014 , 1.749 1015 , 4.048 1014 ]
b_p_values= [0.0 , 0.0 , 0.0 ]
b_distribution= [ student_t , 18 ]
v_estimation= 2.517 10-28
v_conf_int= [1.437 10-28 , 5.504 10-28 ]
v_distribution= [ chi2 , 18 ]
adc= 1.0
    
```

Rys. 6. Wyniki dla macierzy **B** postaci (11), uzyskane poprzez zastosowanie oprogramowania wxMaxima

Drobne różnice wyników z obydwu metod można zaobserwować, gdy pewne wartości macierzy obserwacji spełniają równania (4) z błędami. Przykładowo wprowadzenie drobnego błędu już w jednej obserwacji macierzy **B** (np. $y_9 = 9,79$ zamiast wartości 10,79) daje metodą redukcji zmiennych wyniki przedstawione na rysunku 7.



Rys. 7. Wyniki dla macierzy **B** z drobnym błędem wartości y_j , uzyskane sposobem redukcji zmiennych

Tymczasem stosując oprogramowanie wxMaxima, uzyskuje się w tym przypadku wyniki przedstawione na rysunku 8.

```

LINEAR REGRESSION MODEL
b_estimation= [1.325 , 10.03 , 12.95 ]
b_statistics= [12.7 , 121.3 , 27.9 ]
b_p_values= [2.007 10-10 , 0.0 , 2.22 10-16 ]
b_distribution= [ student_t , 18 ]
v_estimation= 0.05259
v_conf_int= [0.03003 , 0.115 ]
v_distribution= [ chi2 , 18 ]
adc= 0.9987
    
```

Rys. 8. Wyniki dla macierzy **B** z drobnym błędem wartości y_j , uzyskane przy użyciu oprogramowania statystycznego wxMaxima

Porównując wyniki uzyskanych wartości estymatorów b_1 i b_2 z obydwu metod, można zauważyć, że są one obciążone podobnymi błędami procentowymi. Natomiast dużo mniejszy błąd zawiera wartość estymatora b_0 , uzyskana za pomocą sposobu redukcji zmiennych (1,64%) aniżeli otrzymana metodą tradycyjną (5,36%).

Jako kolejny przykład weźmy teraz układ 3D, w którym zmienne x_1, x_2 są zależne, ale w sposób nieliniowy, zaś zmienna zależna y jest postaci:

$$y = 0,6 + 0,8x_1 + 0,4x_2 \quad (12)$$

Dla tego układu wyznaczono macierz obserwacji C postaci:

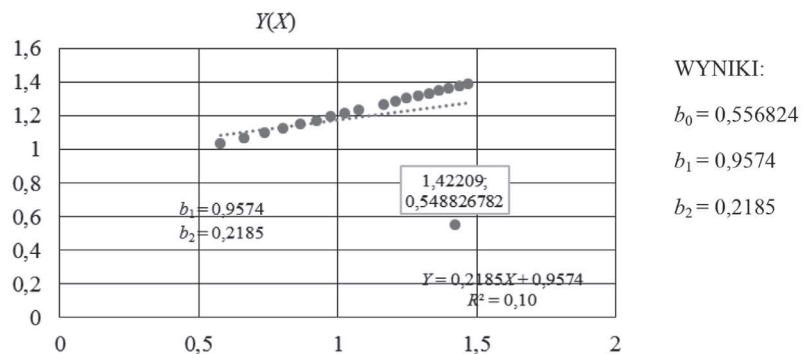
```
C=matrix([0.1,0.031623,0.692649111],[0.2,0.089443,0.795777088],[0.3,0.164317,0.905726707],
[0.4,0.252982,1.021192885],[0.5,0.353553,1.141421356],[0.6,0.464758,1.265903201],
[0.7,0.585662,1.394264807],[0.8,0.715542,1.526216701],[0.9,0.853815,1.661525987],[1.1,1.18],
[1.1,1.45369,1.241475893],[1.2,1.314534,2.085813655],[1.3,1.482228,2.232891221],
[1.4,1.656502,2.382600936],[1.5,1.837117,2.534846923],[1.6,2.023858,2.689543081],
[1.7,2.216529,2.846611527],[1.8,2.414953,3.005981366],[1.9,2.618969,3.167587705], (13)
[2.2,2.828427,3.33137085])
```

Dla powyższej macierzy, stosując profesjonalny program pakietu statystycznego wxMaxima, uzyskano wynik:

```
wyn:linear_regression(C);
apply: second argument must be a list; found: C
#0: linear_regression(dat=C,select=[])(C:\maxima-5.37.2\share\maxima\5.37.2_4_
g0743c29_dirty\share\stats\stats.mac line 1253)
-- an error. To debug this try: debugmode(true);
```

Jak widać, układ danych macierzy C jest zbyt słabo uwarunkowany i profesjonalne oprogramowanie odmówiło podania konkretnych wyników.

Można jednak dla tego układu zastosować sposób redukcji zmiennych opisany w punkcie 2 niniejszej pracy. Stosując ten sposób dla tej samej macierzy C , otrzymano wyniki, które przedstawiono na rysunku 9.



Rys. 9. Wyniki dla macierzy C , uzyskane sposobem redukcji zmiennych

Do obliczeń zastosowano oprogramowanie Excel 2013 w sposób przedstawiony na rysunku 10. Jak łatwo można zauważyć z wykresu $Y(X)$ (rys. 9), jeden wiersz macierzy C ma wadliwe wartości. Wiersz ten odpowiada punktowi o współrzędnych $X = 1,42209$; $Y = 0,548826$, co odpowiada wartości zmiennej $x_1 = 1,1$ (por. rys. 10).

x1	x2	y=0,6+0,8x1+0,4x2	Δx1	Δx2	Δx2/Δx1	Δy/Δx1	b1*x1+b2*x2		średnia z (y-z)
					X	Y	z	y-z	
0,1	0,03162278	0,692649111							
0,2	0,08944272	0,795777088	0,1	0,0578	0,5784272	1,0312798	0,19576603	0,600011056	0,55682406
0,3	0,16431677	0,905726707	0,2	0,1327	0,6635838	1,065388	0,30571314	0,600013568	
0,4	0,25298221	1,021192885	0,3	0,2214	0,7379407	1,0951459	0,42117818	0,600014702	
0,5	0,35355339	1,141421356	0,4	0,322	0,8048835	1,1219306	0,54140671	0,600014645	
0,6	0,464758	1,265903201	0,5	0,4332	0,866316	1,1465082	0,66588968	0,600013524	
0,7	0,58566202	1,394264807	0,6	0,5541	0,9234367	1,1693595	0,79425337	0,600011434	
0,8	0,71554175	1,526216701	0,7	0,6839	0,9770596	1,1908108	0,92620826	0,600008446	
0,9	0,85381497	1,661525987	0,8	0,8222	1,0277687	1,2110961	1,06152137	0,600004619	
1	1	1,8	0,9	0,9684	1,076	1,2303899	1,2	0,6	
1,1	1,45369	1,241475893	1	1,4221	1,42209	0,5488268	1,46151137	-0,220035476	
1,2	1,31453414	2,085813655	1,1	1,2829	1,1663038	1,2665132	1,48582511	0,599988547	
1,3	1,48222805	2,232891221	1,2	1,4506	1,2088567	1,2835351	1,63290944	0,599981777	
1,4	1,65650234	2,382600936	1,3	1,6249	1,2499249	1,2999629	1,78262659	0,59997435	
1,5	1,83711731	2,534846923	1,4	1,8055	1,2896552	1,3158556	1,93488063	0,599966288	
1,6	2,0238577	2,689543081	1,5	1,9923	1,3281718	1,3312626	2,08958547	0,599957614	
1,7	2,21652882	2,846611527	1,6	2,1849	1,3655805	1,3462265	2,24666318	0,599948347	
1,8	2,41495342	3,005981366	1,7	2,3834	1,4019726	1,3607837	2,40604286	0,599938505	
1,9	2,61896926	3,167587705	1,8	2,5874	1,4374274	1,3749659	2,5676596	0,599928103	
2	2,82842712	3,33137085	1,9	2,7968	1,4720143	1,3888009	2,73145369	0,599917157	

Rys. 10. Przedstawienie obliczeń sposobem redukcji zmiennych dla macierzy C postaci (13) (Excel 2013)

Tak więc sposobem redukcji zmiennych łatwo jest wykryć wadliwe „punkty” macierzy obserwacji, co w przypadku wielowymiarowym ($p > 2$) powinno mieć duże znaczenie w praktyce zastosowań.

Łatwo się przekonać, że poprzez usunięcie wadliwej obserwacji (obserwacji dla $x_1 = 1,1$) z macierzy C uzyska się idealne wyniki, korzystając z obydwu sposobów regresji.

4. Zastosowania dla nieliniowego modelu regresji

W przypadku nieliniowego modelu regresji postaci:

$$y = b_0 + b_1 f_1(x_1) + b_2 f_2(x_2) + \dots + b_p f_p(x_p) \quad (14)$$

Gdzie $f_\mu(x_\mu)$ są znanymi funkcjami nieliniowymi dowolnej postaci dla $\mu = 1, 2, \dots, p$, sposób redukcji zmiennych jest identyczny, z tą różnicą, że w równaniu zredukowanym postaci (7) zmienne zastępcze będą równe:

$$Y_i = \frac{\Delta y_i}{\Delta f_{1i}}, \quad X_{vi} = \frac{\Delta f_{vi}}{\Delta f_{1i}}, \quad \delta_i = \frac{\Delta \epsilon_i}{\Delta f_{1i}}, \quad (15)$$

gdzie $v = 2, 3, \dots, p$, zaś

$$\Delta f_{\mu_i} = f_{\mu}(x_{\mu_i}) - f_{\mu}(x_{\mu_1}) \frac{\Delta y_i}{\Delta x_{1i}}. \quad (16)$$

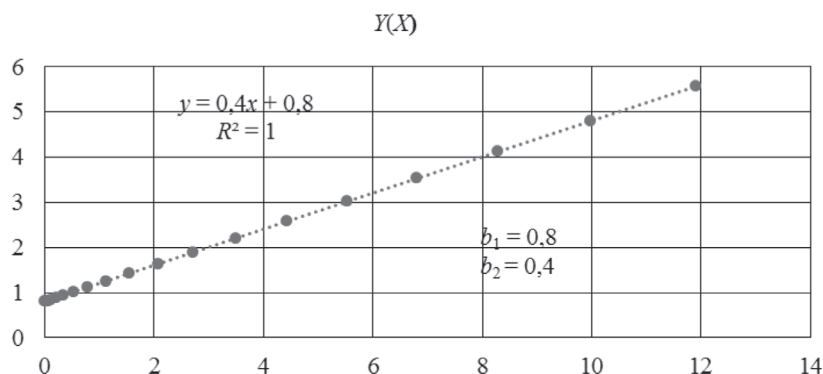
Dla przykładu rozpatrzmy przypadek, w którym zmienna y zależy od x_1, x_2 w sposób określony równaniem:

$$y = 0,6 + 0,8x_1 + 0,4(x_2)^3. \quad (17)$$

Przyjmując te same dane liczbowe zmiennych x_1, x_2 jak w przypadku poprzednim, uzyskuje się sposobem przedstawionym na rysunku 11 idealne wyniki wartości parametrów b_0, b_1, b_2 , które zostały przedstawione na rysunku 12.

x1	x2	y=0,6+0,8x1+0,4(x2)^3	Δx1	(x2)^3	f=Δ[(x2)^3-0,0000316]	Δf/Δx1	Δy/Δx1
						X	Y
0,1	0,0316228	0,680012649		3,16228E-05			
0,2	0,0894427	0,760286217	0,1	0,000715542	0,000683942	0,006839	0,802732
0,3	0,1643168	0,841774621	0,2	0,004436553	0,004404953	0,022025	0,808808
0,4	0,2529822	0,926476345	0,3	0,016190862	0,016159262	0,053864	0,821544
0,5	0,3535534	1,017677767	0,4	0,044194174	0,044162574	0,110406	0,844162
0,6	0,464758	1,120155091	0,5	0,100387728	0,100356128	0,200712	0,880284
0,7	0,585662	1,240352829	0,6	0,200882072	0,200850472	0,334751	0,9339
0,8	0,7155418	1,386542951	0,7	0,366357377	0,366325777	0,523323	1,009329
0,9	0,853815	1,568972445	0,8	0,622431112	0,622399512	0,777999	1,111199
1	1	1,8	0,9	1	0,9999684	1,111076	1,24443
1,1	1,1536897	2,094224414	1	1,535561035	1,535529435	1,535529	1,414211
1,2	1,3145341	2,468605996	1,1	2,27151499	2,27148339	2,064985	1,625994
1,3	1,4822281	2,942582013	1,2	3,256455032	3,256423432	2,713686	1,885474
1,4	1,6565023	3,538176968	1,3	4,545442419	4,545410819	3,49647	2,198588
1,5	1,8371173	4,280108365	1,4	6,200270911	6,200239311	4,428742	2,571497
1,6	2,0238577	5,19588846	1,5	8,289721149	8,289689549	5,52646	3,010584
1,7	2,2165288	6,315922433	1,6	10,88980608	10,88977448	6,806109	3,522443
1,8	2,4149534	7,673603328	1,7	14,08400832	14,08397672	8,284692	4,113877
1,9	2,6189693	9,30540407	1,8	17,96351017	17,96347857	9,97971	4,791884
2	2,8284271	11,2509668	1,9	22,627417	22,6273854	11,90915	5,56366

Rys. 11. Przedstawienie obliczeń metodą redukcji zmiennych dla modelu nieliniowego postaci (17)



Rys. 12. Wyniki obliczeń metodą redukcji zmiennych dla modelu nieliniowego postaci (17)

Jednakże w przypadku gdy kształt funkcji $f_{\mu}(x_{\mu})$ nie jest znany, postuluje się apriorycznie ich liniowość. Dla rozpatrywanego przykładu nowa macierz obserwacji **CN** przyjmie wtedy postać:

$$\mathbf{CN} = \text{matrix}([0.1, 0.031623, 0.680013], [0.2, 0.089443, 0.760286], [0.3, 0.164317, 0.8417746], [0.4, 0.252982, 0.926476], [0.5, 0.353553, 1.017678], [0.6, 0.464758, 1.120155], [0.7, 0.585662, 1.240353], [0.8, 0.715542, 1.386543], [0.9, 0.853815, 1.568972], [1, 1, 1.8], [1.1, 1.15369, 2.094224], [1.2, 1.314534, 2.46806], [1.3, 1.482228, 2.942582], [1.4, 1.656502, 3.538177], [1.5, 1.837117, 4.280108], [1.6, 2.023858, 5.195888], [1.7, 2.216529, 6.315922], [1.8, 2.414953, 7.673603], [1.9, 2.618969, 9.305404], [2, 2.828427, 11.25097]) \quad (18)$$

Stosując metodę analizy regresji dla tej macierzy, otrzymano wyniki, które przedstawiono na rysunku 13.

```

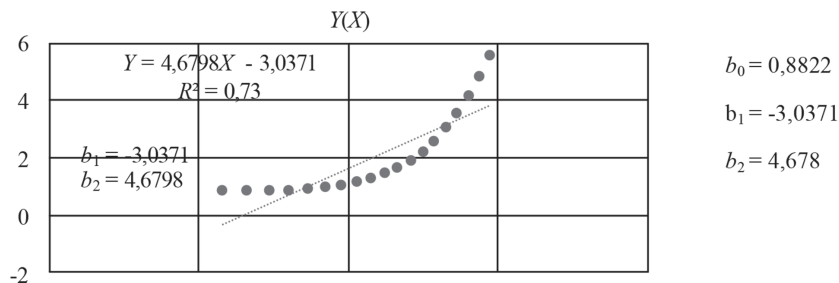
LINEAR REGRESSION MODEL
b_estimation= [ 2.423 , -11.58 , 10.85 ]
b_statistics= [ 4.694 , -6.539 , 9.255 ]
b_p_values= [ 2.089 10-4 , 5.051 10-6 , 4.759 10-8 ]
b_distribution= [ student_t , 17 ]
v_estimation= 0.3703
v_conf_int= [ 0.2085 , 0.8323 ]
v_distribution= [ chi2 , 17 ]
adc= 0.9614
    
```

Rys. 13. Wyniki analizy regresji dla macierzy **CN** (program wxMaxima)

Zauważmy, że chociaż współczynnik korelacji jest bliski jedności (0,9614), to uzyskane wyniki różnią się zdecydowanie od zadanych wartości, zaś otrzymana wartość współczynnika b_1 jest wręcz ujemna.

Innymi słowy, popełnia się duży błąd, przyjmując model liniowy (1) do opisu zależności $y(x_1, x_2)$ w przypadku danych zawartych w macierzy **CN**, o czym użytkownik, nie znając zależności (17), nie ma pojęcia.

Stosując jednak w tym przypadku procedurę przedstawioną w punkcie 2 niniejszej pracy, otrzymujemy wyniki przedstawione na rysunku 14.



Rys. 14. Wyniki dla macierzy **CN** postaci (18), uzyskane metodą redukcji zmiennych

Wyniki te uzyskano w sposób przedstawiony na rysunku 15.

x1	x2	y=0,6+0,8x1+0,4(x2)^3	Δx1	Δx2	Δx2/Δx1	Δy/Δx1	b1*x1+b2*x2	średnia z (y-z)	
					X	Y	z	y-z	
0,1	0,0316228	0,680012649							
0,2	0,0894427	0,760286217	0,1	0,0578427	0,5784272	0,8027322	-0,188846	0,9491322	0,882217
0,3	0,1643168	0,841774621	0,2	0,1327168	0,6635838	0,8088081	-0,1421604	0,983935	
0,4	0,2529822	0,926476345	0,3	0,2213822	0,7379407	0,8215445	-0,0309338	0,9574102	
0,5	0,3535534	1,01767767	0,4	0,3219534	0,8048835	0,8441617	0,1360092	0,8816685	
0,6	0,464758	1,120155091	0,5	0,433158	0,866316	0,8802842	0,3527145	0,7674406	
0,7	0,585662	1,240352829	0,6	0,554062	0,9234367	0,9338997	0,6148111	0,6255417	
0,8	0,7155418	1,386542951	0,7	0,6839418	0,9770596	1,0093285	0,9189123	0,4676307	
0,9	0,853815	1,568972445	0,8	0,822215	1,0277687	1,1111993	1,2622933	0,3066792	
1	1	1,8	0,9	0,9684	1,076	1,24443	1,6427	0,1573	
1,1	1,1536897	2,094224414	1	1,1220897	1,1220897	1,4142114	2,0582272	0,0359972	
1,2	1,3145341	2,468605996	1,1	1,2829341	1,1663038	1,6259936	2,5072369	-0,0386309	
1,3	1,4822281	2,942582013	1,2	1,4506281	1,2088567	1,8854742	2,9883008	-0,0457188	
1,4	1,6565023	3,538176968	1,3	1,6249023	1,2499249	2,1985877	3,5001596	0,0380173	
1,5	1,8371173	4,280108365	1,4	1,8055173	1,2896552	2,5714967	4,0416916	0,2384168	
1,6	2,0238577	5,19588846	1,5	1,9922577	1,3281718	3,0105836	4,6118893	0,5839992	
1,7	2,2165288	6,315922433	1,6	2,1849288	1,3655805	3,5224434	5,2098416	1,1060809	
1,8	2,4149534	7,673603328	1,7	2,3833534	1,4019726	4,1138767	5,834719	1,8388843	
1,9	2,6189693	9,30540407	1,8	2,5873693	1,4374274	4,7918839	6,4857624	2,8196417	
2	2,8284271	11,2509668	1,9	2,7968271	1,4720143	5,5636599	7,1622733	4,0886935	

Rys. 15. Sposób obliczeń wartości modelu regresji dla macierzy CN postaci (18) metodą redukcji zmiennych

Jak widać z wykresu przedstawionego na rysunku 14, funkcja $Y(X)$ nie jest funkcją liniową. Stąd możemy wywnioskować, że zmienna y zależy nieliniowo od zmiennej x_2 , przy czym osobnym problemem pozostaje określenie kształtu tej nieliniowości.

Innymi słowy, uważamy, że zmienna y zależy od zmiennych x_1, x_2 w sposób określony wzorem:

$$f_{\mu}(x_{\mu})y = b_0 + b_1x_1 + b_2f(x_2). \quad (19)$$

Jeśli funkcja $f(x_2)$ byłaby znana, wówczas wartości zmiennej zastępczej X dałoby się wyliczyć ze wzoru:

$$X_i = \frac{f(x_{2i}) - f(x_{21})}{[x_{1i} - x_{11}]}. \quad (20)$$

Związek zmiennej zastępczej Y od tak określonej zmiennej X musi być liniowy – postaci:

$$Y = b_1 + b_2X. \quad (21)$$

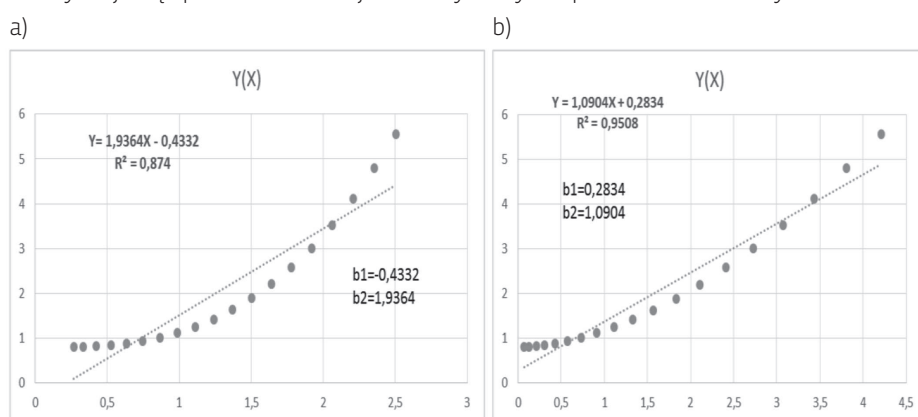
Jeśli w wyniku obliczeń zależność $Y(X)$ jest nieliniowa, wówczas wiadomo, że przyjęta postać funkcji $f(x_2)$ nie jest właściwa. Zauważmy, że występuje to w przykładzie przedstawionym na rysunku 14, gdzie do obliczenia wartości zmiennej X założono $f(x_2) = x_2$. Nie znając prawdziwego kształtu funkcji $f(x_2)$, można kolejno zgadywać, jakiej jest postaci aż do momentu, kiedy zależność $Y(X)$ będzie zależnością liniową. Przykładowo, kolejno zakładając, że funkcja $f(x_2)$ jest być może postaci:

$$f(x_2) = x_2^{1,5} \quad (22)$$

lub też:

$$f(x_2) = x_2^2, \quad (23)$$

uzyskuje się sposobem redukcji zmiennych wyniki przedstawione na rysunku 16.



Rys. 16. Zależności zmiennych zastępczych dla układu (19) przy różnych funkcjach $f(x_2)$: a) – $f(x_2)$ postaci (22), b) – $f(x_2)$ postaci (23)

Porównując wyniki przedstawione na rysunkach 14, 16(a), 16(b), 12, łatwo zauważyć, że w miarę przybliżania się do prawidłowej funkcji $f(x_2) = x^3$ zależność $Y(X)$ staje się coraz bardziej podobna do funkcji liniowej. Rośnie także stopniowo współczynnik korelacji R^2 , by ostatecznie przyjąć wartość $R^2 = 1$ (por. rys. 12). Tak więc, przedstawiony w poprzednim punkcie sposób redukcji zmiennych może być wykorzystywany także, oczywiście w ograniczonym zakresie, do poszukiwania właściwej postaci zależności $y(x_1, x_2, \dots, x_p)$ w sytuacji, gdy zależność ta nie jest liniowa.

5. Przykład zastosowania redukcji zmiennych w przypadku $p > 2$

W niniejszym punkcie przedstawiono przykład zastosowania sposobu redukcji zmiennych w przypadku słabo uwarunkowanego układu liniowego postaci (1) o liczbie $p = 3$. Badania numeryczne przeprowadzono osobno dla trzech różnych macierzy obserwacji różniących się nieznacznie od siebie. Różnice między nimi polegały na wprowadzeniu

drobnej korekty wartości tylko jednej obserwacji i tylko jednej zmiennej. Wybrano zmienną x_2 dla której skorygowano wartość odpowiadającą $i = 11$. Jako wyjściową macierz obserwacji przyjęto macierz P postaci (24), dla której skorygowano jedynie wyraz $p_{11,2}$ czyli wartość 38,6067.

$$P = \begin{bmatrix} 1 & 1 & 0 & 20.3 \\ 3 & 3.7372 & 2.0051 & 23.9181 \\ 5 & 6.8986 & 4.6626 & 27.272 \\ 7 & 10.3304 & 7.6847 & 30.4883 \\ 9 & 13.9666 & 10.9666 & 33.61 \\ 11 & 17.7693 & 14.4527 & 36.6592 \\ 13 & 21.7136 & 18.1081 & 39.6492 \\ 15 & 25.7816 & 21.9086 & 42.5893 \\ 17 & 29.9598 & 25.8367 & 45.4859 \\ 19 & 34.2377 & 29.8788 & 48.344 \\ 21 & 38.6067 & 34.0242 & 51.1675 \\ 23 & 43.0599 & 38.2641 & 53.9595 \\ 25 & 47.5913 & 42.5913 & 56.7226 \\ 27 & 52.1959 & 46.9998 & 59.4589 \\ 29 & 56.8693 & 51.4841 & 62.1703 \\ 31 & 61.6076 & 56.0398 & 64.8584 \\ 33 & 66.4074 & 60.6629 & 67.5245 \\ 35 & 71.2659 & 65.3498 & 70.1699 \\ 37 & 76.1802 & 70.0974 & 72.7956 \\ 39 & 81.1479 & 74.9029 & 75.4026 \end{bmatrix} \quad (24)$$

Ostatnia kolumna tej macierzy została wyliczona dokładnie zgodnie ze wzorem postaci:

$$y = b_0 + b_1x_1 + b_2x_2 + b_3x_3 \quad (25)$$

dla następujących zadanych wartości parametrów b_μ :

$$b_0 = 18, \quad b_1 = 2, \quad b_2 = 0,3, \quad b_3 = -0,6. \quad (26)$$

Wartości liczbowe pozostałych kolumn zostały dobrane tak, aby układ był słabo uwarunkowany. W tym celu przyjęto prawie liniową zależność $x_2(x_1)$ postaci $x_2 = (x_1)^{1,2}$, natomiast wartości zmiennej x_3 obliczono według wzoru:

$$x_3 = x_2 - \sqrt{x_1}. \quad (27)$$

Do obliczeń numerycznych zaokrąglano wszystkie wartości liczbowe do czterech miejsc po przecinku, uzyskując w ten sposób postać (24) macierzy P . Stosując dla tej macierzy typowy program analizy regresyjnej (program wxMaxima), uzyskano w tym przypadku wyniki przedstawione na rysunku 17.

```

                LINEAR REGRESSION MODEL
      b_estimation= [18.0 , 2.0 , 0.3 , -0.6 ]
b_statistics= [2.981 105 , 5.571 104 , 3.669 103 , -8.665 103 ]
      b_p_values= [0.0 , 0.0 , 0.0 , 0.0 ]
      b_distribution= [student_t , 16 ]
      v_estimation= 3.584 10-10
v_conf_int= [1.988 10-10 , 8.301 10-10 ]
      v_distribution= [chi2 , 16 ]
                adc=1.0
    
```

Rys. 17. Wyniki analizy regresji dla macierzy P postaci (24), uzyskane za pomocą programu wxMaxima

Jak widać, otrzymano w tym przypadku wartości idealne równe założonym wartościom (26). Jednakże drobna zmiana tej macierzy może dać już wyniki bardzo różniące się od wartości zadanych (26). Dla przykładu dla macierzy P_p , utworzonej z macierzy P poprzez drobną zmianę wartości wyrazu $p_{11,2} = 38,6067$ na wartość $38,9067$, program wxMaxima dał odpowiedź negatywną (rys. 18), zaś program Scilab 5.5.2 podał wartości estymatorów różniące się znacznie od wartości zadanych (rys. 19).

```

apply: second argument must be a list; found: Pp

#0: linear_regression(dat=Pp,select=[]) (C:\maxima-5.37.2\share\maxima\5.37.2_4_g0743c29_dirty\share\stats\stats.mac line 1253)
-- an error. To debug this try: debugmode(true);
    
```

Rys. 18. Odpowiedź programu wxMaxima dla macierzy P_p

```

[b,bo,a]=reglin(Pp(:,1:3),Pp(:,4))
a = 0.0130185
bo = 18.136673
b =
2.0794108 0.1132107 - 0.4409738
    
```

Rys 19. Odpowiedź programu Scilab 5,5,2 dla macierzy P_p

Dalsza zmiana wartości $p_{11,2}$ macierzy P na wartość $p_{11,2} = 40,6067$ daje za pomocą obydwu programów bardzo podobne wyniki:

$$b_0 = 18,21; \quad b_1 = 2,132; \quad b_2 = -0,00193; \quad b_3 = -0,3446. \quad (28)$$

Zauważmy, że w tym przypadku uległ także zmianie znak estymatora b_2 .

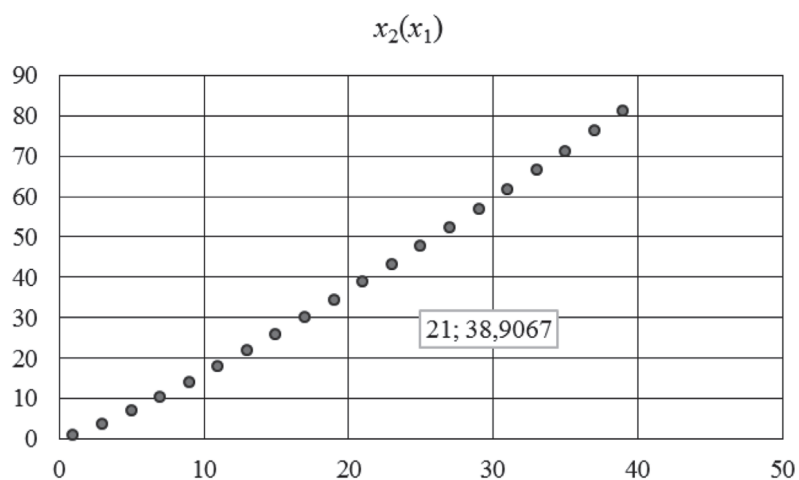
A zatem, stosując profesjonalne oprogramowania analizy regresji w przypadku układów słabo uwarunkowanych, można uzyskać wyniki bardzo różniące się od siebie, w zależności od liczby i zakresów błędnie podanych wartości liczbowych w macierzy obserwacji lub też wartości podawanych z dużym błędem pomiarowym.

Tymczasem, stosując przedstawiony w punkcie 2 niniejszej pracy sposób redukcji zmiennych, uzyskano zarówno dla macierzy P , jak i macierzy Pp i Pp_1 zbliżone wyniki, które zostały przedstawione w tabeli 1.

Tab. 1. Zestawienie wyników uzyskanych sposobem redukcji zmiennych

Macierz	b_0	b_1	b_2	b_3
P	18,0000	2,0000	0,3021	-0,6016
Pp	17,993	2,0052	0,2937	-0,5959
Pp_1	17,966	2,0843	0,1799	-0,5143

Zastosowanie tego sposobu w praktyce przedstawiono poniżej na przykładzie macierzy Pp . W macierzy tej zmienne x_1, x_2 są silnie skorelowane zgodnie z zależnością $x_2 = (x_1)^{1,2}$, z wyjątkiem wyrazu $p_{11'2} = 38,9067$ (rys. 20).



Rys. 20. Wykres zależności $x_2(x_1)$ w macierzy Pp

Zastosowanie sposobu redukcji zmiennych przeprowadzono w tym przykładzie w trzech etapach. W pierwszym etapie zredukowano liczbę zmiennych x_1, x_2, x_3 do dwóch zmiennych zastępczych X_2, X_3 (rys. 21).

x1	x2 (zaokr B)	x3 (zaokr C)	y (zaokr D)	$\Delta x1$	$\Delta x2$	$\Delta x3$	Δy	X2	X3	Y
1	1	0	20,3					$\Delta x2/\Delta x1$	$\Delta x3/\Delta x1$	$\Delta y/\Delta x1$
3	3,7372	2,0051	23,9181	2	2,7372	2,0051	3,6181	1,3686	1,00255	1,80905
5	6,8986	4,6626	27,272	4	5,8986	4,6626	6,972	1,47465	1,16565	1,743
7	10,3304	7,6847	30,4883	6	9,3304	7,6847	10,188	1,55506667	1,280783	1,69805
9	13,9666	10,9666	33,61	8	12,9666	10,9666	13,31	1,620825	1,370825	1,66375
11	17,7693	14,4527	36,6592	10	16,7693	14,4527	16,359	1,67693	1,44527	1,63592
13	21,7136	18,1081	39,6492	12	20,7136	18,1081	19,349	1,72613333	1,509008	1,61243
15	25,7816	21,9086	42,5893	14	24,7816	21,9086	22,289	1,77011429	1,5649	1,59209
17	29,9598	25,8367	45,4859	16	28,9598	25,8367	25,186	1,8099875	1,614794	1,57412
19	34,2377	29,8788	48,344	18	33,2377	29,8788	28,044	1,84653889	1,659933	1,558
21	38,9067	34,0242	51,1675	20	37,9067	34,0242	30,868	1,895335	1,70121	1,54338
23	43,0599	38,2641	53,9595	22	42,0599	38,2641	33,66	1,91181364	1,739277	1,52998
25	47,5913	42,5913	56,7226	24	46,5913	42,5913	36,423	1,94130417	1,774638	1,51761
27	52,1959	46,9998	59,4589	26	51,1959	46,9998	39,159	1,96907308	1,807685	1,50611
29	56,8693	51,4841	62,1703	28	55,8693	51,4841	41,87	1,99533214	1,838718	1,49537
31	61,6076	56,0398	64,8584	30	60,6076	56,0398	44,558	2,02025333	1,867993	1,48528
33	66,4074	60,6629	67,5245	32	65,4074	60,6629	47,225	2,04398125	1,895716	1,47577
35	71,2659	65,3498	70,1699	34	70,2659	65,3498	49,87	2,06664412	1,922053	1,46676
37	76,1802	70,0974	72,7956	36	75,1802	70,0974	52,496	2,08833889	1,94715	1,45821
39	81,1479	74,9029	75,4026	38	80,1479	74,9029	55,103	2,10915526	1,971129	1,45007

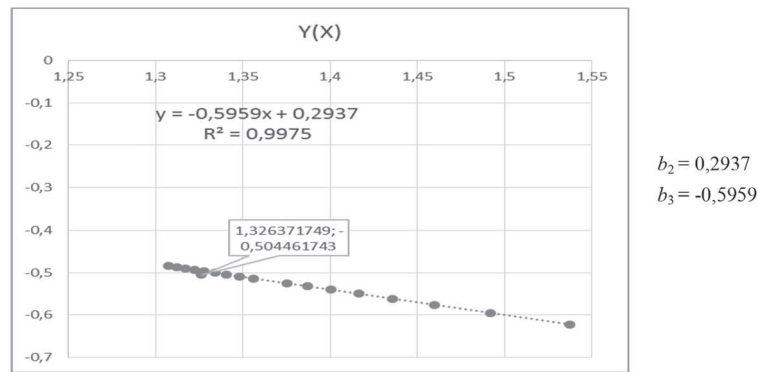
Rys. 21. Obliczenia pierwszego etapu redukcji zmiennych dla macierzy Pp (Excel 2013)

Następnie w etapie drugim powtórzono procedurę dla dwóch zmiennych X_2, X_3 , redukując liczbę zmiennych niezależnych do jednej zmiennej zastępczej X i jednej zmiennej zależnej Y (rys. 22).

X2	X3	Y	x1	x2	y	$\Delta x1$	$\Delta x2$	Δy	X	Y
x1	x2	y	zaokr x1	zaokr x2	zaokr y				$\Delta x2/\Delta x1$	$\Delta y/\Delta x1$
1,3686	1,00255	1,80905	1,3686	1,0026	1,8091					
1,47465	1,16565	1,743	1,4747	1,1657	1,743	0,1061	0,1631	-0,066	1,53723	-0,623
1,555067	1,280783	1,69805	1,5551	1,2808	1,6981	0,1865	0,2782	-0,111	1,49169	-0,59517
1,620825	1,370825	1,66375	1,6208	1,3708	1,6638	0,2522	0,3682	-0,145	1,45995	-0,57613
1,67693	1,44527	1,63592	1,6769	1,4453	1,6359	0,3083	0,4427	-0,173	1,43594	-0,56179
1,726133	1,509008	1,612433	1,7261	1,509	1,6124	0,3575	0,5064	-0,197	1,4165	-0,55021
1,770114	1,5649	1,592093	1,7701	1,5649	1,5921	0,4015	0,5623	-0,217	1,4005	-0,54047
1,809988	1,614794	1,574119	1,81	1,6148	1,5741	0,4414	0,6122	-0,235	1,38695	-0,5324
1,846539	1,659933	1,558	1,8465	1,6599	1,558	0,4779	0,6573	-0,251	1,37539	-0,52542
1,895335	1,70121	1,543375	1,8953	1,7012	1,5434	0,5267	0,6986	-0,266	1,32637	-0,50446
1,911814	1,739277	1,529977	1,9118	1,7393	1,53	0,5432	0,7367	-0,279	1,35622	-0,51381
1,941304	1,774638	1,517608	1,9413	1,7746	1,5176	0,5727	0,772	-0,292	1,348	-0,50899
1,969073	1,807685	1,506112	1,9691	1,8077	1,5061	0,6005	0,8051	-0,303	1,34072	-0,50458
1,995332	1,838718	1,495368	1,9953	1,8387	1,4954	0,6267	0,8361	-0,314	1,33413	-0,50056
2,020253	1,867993	1,48528	2,0203	1,868	1,4853	0,6517	0,8654	-0,324	1,32791	-0,49685
2,043981	1,895716	1,475766	2,044	1,8957	1,4758	0,6754	0,8931	-0,333	1,32233	-0,49349
2,066644	1,922053	1,466762	2,0666	1,9221	1,4668	0,698	0,9195	-0,342	1,31734	-0,4904
2,088339	1,94715	1,458211	2,0883	1,9472	1,4582	0,7197	0,9446	-0,351	1,31249	-0,48756
2,109155	1,971129	1,450068	2,1092	1,9711	1,4501	0,7406	0,9685	-0,359	1,30772	-0,48474

Rys. 22. Obliczenia drugiego etapu redukcji zmiennych dla macierzy Pp (Excel 2013)

Tak uzyskane wartości zmiennych X, Y posłużyły bezpośrednio do obliczenia estymatorów b_2, b_3 . W tym celu zastosowano prostą regresję programu Excel danych liczbowych tych zmiennych przedstawionych na wykresie $Y(X)$ (rys. 23).



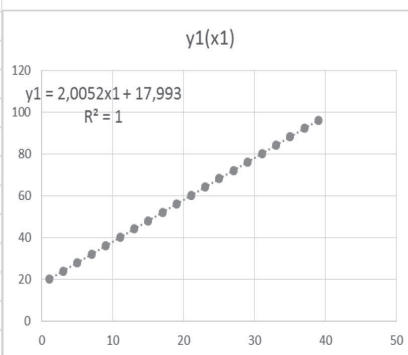
Rys. 23. Sposób obliczenia wartości b_2, b_3 dla danych macierzy Pp

Następnie, wprowadzając nową zmienną zależną y_1 postaci:

$$y_1 = y - (0,2937x_2 - 0,5959x_3) \quad (29)$$

przeprowadzono kolejną prostą regresję zależności $y_1(x_1)$, uzyskując w ten sposób wartości b_0, b_1 podane w tabeli 1 (rys. 24).

x1	x2 (zaokr B)	x3 (zaokr C)	y (zaokr D)	$z=0,2937x_2-0,5959x_3$	$y_1=y-z$
1	1	0	20,3	0,2937	20,0063
3	3,7372	2,0051	23,9181	-0,09722345	24,01532
5	6,8986	4,6626	27,272	-0,75232452	28,02432
7	10,3304	7,6847	30,4883	-1,54527425	32,03357
9	13,9666	10,9666	33,61	-2,43300652	36,04301
11	17,7693	14,4527	36,6592	-3,39352052	40,05272
13	21,7136	18,1081	39,6492	-4,41333247	44,06253
15	25,7816	21,9086	42,5893	-5,48327882	48,07258
17	29,9598	25,8367	45,4859	-6,59689627	52,0828
19	34,2377	29,8788	48,344	-7,74916443	56,09316
21	38,9067	34,0242	51,1675	-8,84812299	60,01562
23	43,0599	38,2641	53,9595	-10,15488456	64,11438
25	47,5913	42,5913	56,7226	-11,40259086	68,12519
27	52,1959	46,9998	59,4589	-12,67724499	72,13614
29	56,8693	51,4841	62,1703	-13,97686178	76,14716
31	61,6076	56,0398	64,8584	-15,2999647	80,15836
33	66,4074	60,6629	67,5245	-16,64516873	84,16967
35	71,2659	65,3498	70,1699	-18,01115099	88,18105
37	76,1802	70,0974	72,7956	-19,39691592	92,19252
39	81,1479	74,9029	75,4026	-20,80149988	96,2041



Rys. 24. Sposób obliczenia wartości estymatorów b_0, b_1 dla danych macierzy Pp

Identyczny tok obliczeń wykonano dla macierzy \mathbf{P} i Pp_{μ} , uzyskując wyniki podane w tabeli 1.

6. Podsumowanie i wnioski

Przedstawiony w punkcie 2 niniejszej pracy sposób obliczeń parametrów modelu regresji daje dobre wyniki zarówno dla układów dobrze, jak i słabo uwarunkowanych. Ponadto, dzięki zastąpieniu poszukiwania optymalnych parametrów modelu regresji w przestrzeni $p + 1$ -wymiarowej przez poszukiwanie dwu wybranych parametrów w przestrzeni D_2 , bardzo łatwo jest w sposób wizualny określić wadliwe wyrazy macierzy obserwacji, a także sprawdzić, czy spełniony jest postulat liniowości modelu regresji.

Osobnym problemem pozostaje sposób wyboru tej obserwacji, która ma być traktowana jako pierwsza ($i = 1$) i odejmowana jest od pozostałych. Ponieważ numeracja obserwacji jest dowolna, jako pierwszą należy przyjmować tę, która jest najbardziej wiarygodna (np. mierzona wielokrotnie dla tych samych wartości zmiennych niezależnych). Ponadto różnice Δx_{μ} muszą być istotnie różne od zera, aby dało się przedstawić metodę swobodnie stosować. Dalsze badania w tym zakresie są planowane. Problemem pozostaje także wyznaczanie rozkładów prawdopodobieństwa i statystyk wiarygodności wyznaczanych w ten sposób parametrów estymatorów b_{μ} .

Literatura

1. Mańczak K., *Metody identyfikacji wielowymiarowych obiektów sterowania*, Wydawnictwo WNT, Warszawa, 1971.
2. Johnston J. J., *Econometric Methods*, McGraw-Hill Book Company Inc, New York 1963.
3. Hill P. D. H., *A review of experimental design procedures for regression model discrimination*, „Technometrics”, nr 10, 1978, s. 145–160.
4. Draper N. R., Smith H., *Applied regression analysis*, 2nd ed., Wiley, New York 1981.
5. Fox J., *Regression diagnostics*, Sage, Newbury Park 1977.
6. Mosteller F., Tukey J. W., *Data analysis and regression: A second course in statistics*, Addison-Wesley, Reading 1977.
7. Neter J., Kutner M. H., Nachtsheim C. J., Wasserman W., *Applied linear statistical models*, 4th ed., IRWIN, Toronto 1996.
8. Bard Y., *Nonlinear parameter estimation*, Academic Press, New York 1974.
9. Gallant A. R., *Nonlinear statistical models*, Wiley, New York 1987.
10. Hunter J. J., *Mathematical techniques of applied Probability*, t. 1: *Discrete Time Models: Basic Theory*, Academic Press, New York 1983.
11. Kendall M. G., Stuart A., *The advanced theory of statistics*, t. 2, Griffin, London 1973.
12. Kennedy W. J. J., Gentle J. E., *Statistical Computing*, Marcel Dekker, New York 1980.
13. Maindonald J. H., *Statistical Computation*, Wiley, New York 1984.

14. Polak E., *Computational Methods in Optimization: A Unified Approach*, Academic Press, New York 1973.
15. Snedecor G. W., Cochran W. G., *Statistical Methods*, Iowa State University Press, Ames 1967.
16. Spendley W., *Nonlinear least squares fitting using a modified simplex minimization method*, [w:] R. Fletcher (red.), *Optimization*, Academic Press, London, New York 1969, s. 259–270.
17. Rhinehart R. R., *Nonlinear Regression Modeling for Engineering Applications: Modeling, Model Validation, and Enabling Design of Experiments*, Wiley, New York 2016.