Andrzej KORNACKI, Mirosława WESOŁOWSKA-JANCZAREK
Katedra Zastosowań Matematyki i Informatyki
Uniwersytet Przyrodniczy
ul. Akademicka 13, 20-950 Lublin
e-mail: andrzej.kornacki@up.lublin.pl

COMPARING OF COEFFICIENTS R^2 AND *RMS* USED IN VERIFICATIONS OF CORRECTNESS MATHEMATICAL MODELS ON BASIC ON EXPERIMENTAL DATA

Summary

R^2 and RMS are used for verification of correctness of fitting models. In this paper we try to test if both coefficients give the same conclusions about goodness of fit of model to experimental data, and which of them may be recommended to practicians. Research were conducted by means of computer simulations with four kinds of functions, two kinds of distributions of x: normal and uniform and three ranges values of x.

PORÓWNANIE WSPÓŁCZYNNIKÓW R^2 I *RMS* UŻYWANYCH PRZY WERYFIKOWANIU POPRAWNOŚCI MODELI MATEMATYCZNYCH W OPARCIU O DANE EMPIRYCZNE

Streszczenie

Wśród współczynników wykorzystywanych do weryfikowania poprawności dopasowanych modeli są R^2 i RMS. W niniejszej pracy podjęto próbę sprawdzenia, czy oba te współczynniki dają zgodne wnioski o dobroci dopasowania modelu do danych empirycznych i wskazania, który z tych współczynników można byłoby polecać praktykom. Badania przeprowadzono metodą symulacji komputerowej przy czterech różnych typach funkcji, dwóch różnych rozkładach x: normalnym i jednostajnym oraz trzech różnych zakresach zmian wartości x.

1. Wstęp

Badanie dobroci dopasowania modeli regresyjnych do danych empirycznych jest jednym z częściej rozważanych problemów w analizie wyników doświadczeń. Jedną z powszechnie stosowanych miar jest współczynnik determinacji R^2.

Współczynnik R^2 jest miarą dopasowania, nie zawsze najlepszą. Stąd też w literaturze statystycznej można znaleźć prace, w których prezentowane są nowe sposoby wyznaczania tego współczynnika i badane ich własności (np.: Barrett [1], Rencher i Pun [6], Magge [4] czy Nagelkere [5] lub proponowane są inne nowe współczynniki dopasowania, jak współczynnik *d* w pracy Kornackiego i Wesołowskiej-Janczarek [3]. Przegląd tych współczynników wraz z przykładem ich zastosowania do danych empirycznych można znaleźć w pracy Kornackiego i Wesołowskiej-Janczarek [3]. Uwzględnione tam współczynniki nie zawsze dają jednoznaczną odpowiedź, co do dobroci dopasowania modelu. Nie można jednak wskazać konkretnych kryteriów ich wykorzystania (Magge [4]).

W pracach z inżynierii rolniczej można jeszcze znaleźć kolejny współczynnik, oznaczony *RMS*, używany jako miara dobroci dopasowania modelu do danych empirycznych (Białobrzeski [2]).

W tej pracy porównujemy współczynniki R^2 i *RMS* na podstawie danych symulowanych przy wykorzystaniu arkusza kalkulacyjnego EXCEL. Kolejne części pracy zawierają: krótkie przypomnienie współczynników, opis przyjętej metody symulacji danych, uzyskane wyniki estymacji parametrów w przyjętych modelach i obliczone wartości R^2 i *RMS* oraz podsumowanie i wnioski.

2. Porównywane współczynniki

Wybrane do porównania współczynniki to zwykły współczynnik determinacji R^2 oraz (*RMS*) (*Root Mean Square*). Przypomnijmy je krótko:

a) Współczynnik R^2 dla cechy *y* definiowany jest jako

$$R^2 = \frac{SS_{regr}}{SS_y} = 1 - \frac{SS_e}{SS_y}, \qquad (1)$$

gdzie $SS_y = \sum_{i=1}^{n} (y_i - \bar{y})^2$ jest sumą kwadratów odchyleń od średniej dla badanej cechy, $SS_{regr} = \sum_{i=1}^{n} (Y_i - \bar{y})^2$ jest sumą kwadratów dla regresji; $SS_e = \sum_{i=1}^{n} (y_i - Y_i)^2$ jest sumą kwadratów odchyleń od regresji; y_i $(i = 1,...,n)$ są wartościami obserwowanymi a Y_i wartościami dopasowanymi z oszacowanego równania, natomiast *n* jest liczebnością próby.

Współczynnik ten dla zwykłej regresji prostoliniowej jest równy kwadratowi współczynnika korelacji r_{xy}^2 między badanymi cechami. Przyjmuje on na ogół wartości z przedziału $\langle 0,1 \rangle$. Obliczone ujemnie wartości współczynnika determinacji R^2 mogą wskazywać na źle wybraną funkcję regresji. Nie powinien być on też używany do badania dopasowania logistycznej regresji (Ryan [7, str. 447] wskazuje też na możliwość pojawienia się ujemnej wartości R^2 obliczanej według wzoru (2.1), gdy regresja nie zawiera wyrazu wolnego. Obliczona wartość współczynnika R^2 bliższa 1 wskazuje na lepsze dopasowanie krzywej, a bliższa 0 na złe dopasowanie.

b) Współczynnik *RMS* (*Root Mean Square*) wyraża się wzorem:

$$RMS = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\frac{y_i - Y_i}{y_i}\right)^2},$$

gdzie jak poprzednio y_i są wartościami empirycznymi, zaś Y_i – wartościami oszacowanymi z równania, a *n*-liczbą obserwacji. Współczynnik ten przyjmuje wartości dodatnie. Wartość współczynnika bliższa 0 wskazuje na lepsze dopasowanie oszacowanej krzywej do punktów empirycznych, ale jak łatwo można zobaczyć w tabelach zamieszczonych poniżej, może przyjmować dowolnie duże wartości, co nie ułatwia interpretacji obliczonej wartości. W dalszej części pracy zostaną porównane wyniki uzyskiwanych wartości obu tych współczynników.

3. Opis zastosowanej metody symulacyjnej

Dla wybranych dowolnie funkcji czterech typów: wielomianu, potęgowej, logarytmicznej i wykładniczej przy założeniu dla zmiennej niezależnej *x* rozkładu normalnego o określonych parametrach μ i σ, co zapisujemy $N(\mu,\sigma)$ lub jednostajnego określonego na przedziale o odpowiedniej długości, czyli przedziału (a,b) oznaczonego $J(a,b)$ generowano losowo po 200 wartości *x* i obliczono *y*, a następnie wyznaczono „dopasowywaną" krzywą opcją „dopasuj linie trendu". Rozpatrzono przypadki, w których obie krzywe dana i dopasowana były wielomianami, pierwsza z nich była wielomianem a dopasowana każdego z rozważanych innych typów funkcji pierwsza z nich była innego typu, a dopasowywana wielomianem, a wreszcie zarówno dana jak i dopasowana była innego typu niż wielomian.

Dla każdej z tych 16 kombinacji uwzględniono również różny zakres zmian wartości *x*. Dla rozkładu normalnego w zakresie $(\mu - 3\sigma, \mu + 3\sigma)$ dla $\sigma = 1; 5; 10$ i różnych μ, a dla rozkładu jednostajnego podobnie ustalono długość zakresu wartości *x* na 6, 30 i 60. Każda kombinacja została powtórzona 4 razy, zatem rozważono 96 możliwości. Uporządkowane wyniki obejmujące krzywe dane, dopasowane oraz obliczone współczynniki R^2 i *RMS* oraz ich porównanie są przedstawione w dalszej części pracy.

4. Analiza wyników badań

Jak już wspomniano wcześniej na lepsze dopasowanie modelu wskazuje wartość R^2 bliższa 1, ale równocześnie mniejsza, bliższa zeru, wartość *RMS*. Obliczona wartość współczynnika korelacji dla obliczonych tych współczynników jest rzeczywiście ujemna, ale korelacja jest bardzo słaba, $r = -0{,}075$, co wskazuje na bardzo słabą zgodność wniosków wynikających z interpretacji tych wskaźników. Największą wartością współczynnika *RMS* jest $3 \cdot 10^8$ przy dopasowaniu funkcji potęgowej, gdy danym był wielomian i wartości *x* były losowane z rozkładu jednostajnego z przedziału (3; 63), podczas gdy wartość współczynnika R^2 była dość duża co mogłoby świadczyć o nienajgorszym dopasowaniu (pozycja 90 w tab. 1). Największa zmienność wśród wartości *RMS* widoczna jest, gdy zadaną była funkcja potęgowa a dopasowywany był wielomian (poz. 76 i 90) i kolejno dana była funkcja potęgowa dopasowana logarytmicznie (poz. 84 i 95), dany wielomian, a dopasowaną funkcja logarytmiczna (poz. 5 i 29), daną funkcja wykładnicza – dopasowaną wielomian (poz. 53 i 69) i dana potęgowa a dopasowana wykładnicza (poz. 79 i 92). Współczynnik R^2 nie wykazywał w tych samych przypadkach silnego zróżnicowania.

Warto jeszcze zwrócić uwagę na to, że te największe wartości *RMS* wcale nie wszędzie odpowiadały najmniejszym wartościom R^2 i odwrotnie. Jako przykład można wskazać $R^2 = 0{,}9967$ i $RMS = 5465{,}7$ (poz. 92 tab. 1), czy też $R^2 = 8 \cdot 10^{-5}$ podczas, gdy $RMS = 1{,}1670$ (poz. 84 tab. 1) co pociągnęłoby za sobą zdecydowanie przeciwne wnioski. Podobnie różniące się wnioski co do dobrego dopasowania funkcji po obliczeniu R^2 i *RMS*, mogłyby być sformułowane w przypadku dopasowywania wielomianu do funkcji potęgowej czy wielomianu do funkcji wykładniczej bez względu na to z jakiego rozkładu pochodziły wartości *x*.

W pozostałych przypadkach różnice nie są, aż tak zdecydowane, ale też mogłyby być podstawą do rozbieżnych wniosków. Przykładem mogą tu być chociażby takie, gdy dopasowany wielomian do danej funkcji logarytmicznej wykazał raz $R^2 = 0{,}1941$ a $RMS = 11{,}5380$ dla rozkładu normalnego, podczas gdy $R^2 = 0{,}6633$, a $RMS = 10{,}428$ (poz. 44 tab. 1) przy rozkładzie jednostajnym, i $R^2 = 0{,}1394$ a $RMS = 5{,}7348$ (poz. 20 tab. 1) dla rozkładu normalnego. Nie widać więc żadnej regularności zmian tych współczynników. Warto jeszcze zwrócić uwagę na przypadki, w których oba współczynniki są prawie równe, jak np. $R^2 = 0{,}2578$ i $RMS = 0{,}2581$, gdy dopasowana była funkcja wykładnicza do potęgowej (poz. 63 tab. 1) oraz $R^2 = 0{,}1930$, a $RMS = 0{,}1462$, gdy dopasowana była funkcja wykładnicza, podczas gdy dana była logarytmiczna (poz. 18 tab. 1). W pierwszym z tych przypadków rozkład *x* był jednostajny, a w drugim normalny.

Warto również zwrócić uwagę jak zmieniają się współczynniki R^2 i *RMS* przy odpowiednich rozkładach i zakresach zmian dla wartości *x*, bez względu na kształt funkcji danej i dopasowywanej. Odpowiednie wyniki są zamieszczone w tab. 3.

Z tab. 3 widać, że największą zgodność wniosków i najmniejsze zróżnicowanie *RMS* można zaobserwować w przypadku, gdy *x* były losowane z rozkładu normalnego o małym $\sigma = 1$ czyli o małym zróżnicowaniu wartości *x* (potwierdza to wartość $r = 0{,}076$ (tab. 4)), chociaż i tu zdarzają się takie sytuacje jak $R^2 = 0{,}7986$ i $RMS = 10{,}01$, $R^2 = 0{,}7639$ i $RMS = 73{,}169$ oraz $R^2 = 0{,}9008$ i $RMS = 21{,}759$. W przypadku rozkładu jednostajnego wartości *RMS* są silnie zróżnicowane, ale daje się zauważyć nieco większa zbieżność wniosków (tab. 3) przy najwyższej korelacji ujemnej $r = -0{,}081$ (tab. 4).

Tab. 1. Funkcja dana, dopasowana, R^2, *RMS* i rozkład *X*
Table 1. Given function, fitting function, R^2, RMS and distribution of X

Lp.	Dana funkcja	Dopasowana funkcja	R^2	*RMS*	*Rozkład x*
1	y=x^2-3x+2	y=1,0522x^2-3,5865x+3,5308	0,9818	0,1679	N(5,1)
2	y=2x^2+4x-2	y=1,996x^2+4,0228x-2,1044	0,9999	0,794	N(4,5)
3	y=x^3-3x^2+2	y=x^3-3x^2+0,0112x+1,8101	1	0,8147	N(3,10)
4	y=x^2-3x+2	y=15,946ln(x)-14,314	0,7639	73,169	N(4,1)
5	y=3x^2-4x+5	y=4551,4ln(x)-12,781	0,9475	0,1458	N(30,5)
6	y=2x^3-3x^2+4x-5	y=589785ln(x)-2000000	0,6237	55,597	N(50,10)
7	y=x^2-4x+5	y=7,4968exp(0,4017x)	0,9445	0,0686	N(3,1)
8	y=2x^3-x^2+3x-5	y=21,079exp(0,203x)	0,8665	1,2432	N(4,5)
9	y=3x^2-5x+6	y=78,132exp(0,0999x)	0,6547	5,8677	N(8,10)
10	y=x^2+2x-4	y=1,1817x^(2,0212)	0,9914	0,0438	N(5,1)
11	y=2x^3+x^2+5x+100	y=3,5091x^(2,8419)	0,9991	0,0207	N(7,5)
12	y=2x^2+3x+5	y=2,3831x^(1,9633)	1	0,0016	N(50,10)
13	y=2,35ln(x)-3,12	y=2,4618ln(x)-2,8298	0,7986	10,01	N(5,1)
14	y=2,1ln(x)+3	y=2,0306ln(x)+3,019	0,2715	0,1109	N(20,5)
15	y=-1,5ln(x)+2	y=-1,3742ln(x)+1,3502	0,1256	0,5349	N(40,10)
16	y=2,2ln(x)-3	y=1,4206exp(0,3498x)	0,8411	0,2281	N(4,1)
17	y=3,2ln(x)+1	y=8,1227exp(0,0077x)	0,1237	0,0986	N(30,5)
18	y=1,5ln(x)+2	y=5,5115exp(0,0067x)	0,193	0,1462	N(40,10)
19	y=2,1ln(x)-2	y=-0,1482x^2+1,8753x-4,3671	0,1941	11,538	N(5,1)
20	y=-1,6ln(x)+3	y=0,0012x^2-0,1263x+0,0894	0,1394	5,7348	N(20,5)
21	y=3,2ln(x)+1	y=0,0014x^2+0,3005x+3,1308	0,9971	0,009	N(40,10)
22	y=2,3ln(x)+3,1	y=3,2623x^(0,4389)	0,2815	0,1914	N(4,1)
23	y=1,4ln(x)+0,9	y=1,5921x^(0,3587)	0,0986	0,2051	N(30,5)
24	y=-1,2ln(x)+3,1	y=19,035x^(-0,5623)	0,0692	6,1843	N(40,10)
25	y=4x^3-2x^2+x-5	y=9,89x^3-11,57x^2+4,48x-5,04	1	0,4805	J(0,6)
26	y=4x^2+2x-5	y=3,9547x^2+2,7099x-7,8714	1	0,0024	J(5,35)
27	y=2x^2+3x+5	y=1,986x^2+3,1991x+4,259	1	0,008	J(2,62)
28	y=2x^3-x^2+3x+4	y=316,06ln(x)-212,94	0,6954	19,091	J(1,7)
29	y=3x^3-x^2+2x-1	y=14657ln(x)-15067	0,4449	7904,6	J(0,30)
30	y=3x^2+2x-1	y=4308,9ln(x)-10030	0,8855	0,7284	J(3,63)
31	y=2x^2+4x-3	y=8,2829exp(0,3921x)	0,9531	0,1226	J(2,8)
32	y=3x^3+x^2-2x+4	y=54,87exp(0,2914x)	-4,685	3,207	J(0,30)
33	y=4x^2+2x-3	y=250,22exp(0,0748x)	0,6059	0,6641	J(4,64)
34	y=2x^2+4x-3	y=3,644x^(1,8094)	0,998	0,0306	J(2,8)
35	y=3x^3+x^2-2x+4	y=12,529x^(2,4782)	0,8786	0,4602	J(0,30)
36	y=4x^2+2x-3	y=4,3959x^(1,9778)	1	0,0052	J(4,64)
37	y=2,35ln(x)+3,12	y=2,3673ln(x)+3,1798	0,863	0,5346	J(0,6)
38	y=2,1ln(x)+3	y=2,0573ln(x)+3,1982	0,5389	0,1135	J(5,35)
39	y=-1,5ln(x)+2	y=-1,4339ln(x)+1,8603	0,5575	29,31	J(2,62)
40	y=2,2ln(x)-3	y=7,4147exp(-0,2385x)	0,4645	0,9768	J(1,7)
41	y=3,2ln(x)+1	y=12,126exp(-0,0506x)	0,7676	0,2896	J(0,30)
42	y=1,5ln(x)+2	y=4,9886exp(0,0096x)	0,456	0,2027	J(3,63)
43	y=2,1ln(x)-2	y=-0,02x^4+0,4x^3-2,5x^2+8,3x-9,8	0,461	33,367	J(2,8)
44	y=-1,6ln(x)+3	y=0,0091x^2-0,4305x+3,0183	0,6633	10,428	J(0,30)
45	y=3,2ln(x)+1	y=-0,0023x^2+0,2773x+5,7106	0,826	0,1124	J(4,64)
46	y=2,3ln(x)+3,1	y=y=3,7134exp(0,3709)	0,3431	2,285	J(2,8)
47	y=1,4ln(x)+0,9	y=2,4141x^(0,3491)	0,5794	0,2942	J(0,30)
48	y=-1,2ln(x)+3,1	y=11,876x^(-0,4094)	0,2798	7,686	J(4,64)
49	y=2,4exp(1,5x)	y=2,3978exp(1,5002x)	1	0,0023	N(5,1)
50	y=2,8exp(0,0011x)	y=2,87exp(0,00002x)	0,0003	1,1549	N(4,5)
51	y=3,5exp(0,01x)	y=3,3898exp(0,012x)	0,0643	0,7863	N(3,10)
52	y=2,4exp(1,5x)	y=1313,7x^2-7580x+10420	0,9008	21,759	N(4,1)
53	y=3,2exp(2,7x)	y=0,144x^2-5,9256x+71,928	0,9916	0,0889	N(30,5)
54	y=1,1exp(0,2x)	y=3,491x^2-282,93x+5258,1	0,7782	138,68	N(50,10)
55	y=2,4exp(1,5x)	y=1277,9ln(x)-681,04	0,1865	995,11	N(3,1)
56	y=3,2exp(2,7x)	y=7503,4ln(x)-20320	0,1199	173,82	N(4,5)
57	y=0,7exp(0,25x)	y=624016ln(x)-2*10^6	0,126	5702,7	N(35,10)
58	y=2,4exp(1,5x)	y=0,0551x^(7,094)	0,7019	0,1641	N(5,1)
59	y=1,2exp(0,5x)	y=2*10^(-7)*x^8,7123	0,5798	0,5213	N(20,5)
60	y=1,5exp(3,2x)	y=4*10^(-12)*x^9,3464	0,7505	0,2427	N(50,10)
61	y=2,4exp(1,5x)	y=38,362x^(2,2483)	0,1081	1,5978	J(0,6)
62	y=2,8exp(0,8x)	y=9*10^(-10)*x^(13,118)	0,028	2,6504	J(5,35)
63	y=3,5exp(0,01x)	y=2,8855x^(0,1584)	0,2578	0,2581	J(2,62)

c.d. Tab. 1.

Lp.	Dana funkcja	Dopasowana funkcja	R^2	*RMS*	*Rozkład x*
64	y=2,4exp(1,5x)	y=2,3931exp(1,5006x)	1	0,0276	J(1,7)
65	y=3,2exp(2,7x)	y=3,0764exp(2,7019x)	0,9997	0,916	J(0,30)
66	y=1,1exp(0,2x)	y=1,0622exp(0,2009x)	0,9996	0,2398	J(3,63)
67	y=2,4exp(1,5x)	y=6760x^3-84090x^2+331036x-408136	0,9621	75,803	J(2,8)
68	y=3,2exp(2,7x)	y=0,9042x^2-16,013x+50,428	0,9336	136,82	J(0,30)
69	y=0,7exp(0,25x)	y=2361,4x^2-123490x+10^6	0,668	82216	J(4,64)
70	y=2,4exp(1,5x)	y=116317ln(x)-138650	0,4124	220	J(2,8)
71	y=1,2exp(0,5x)	y=217865ln(x)-240937	0,0572	419942	J(0,30)
72	y=1,2exp(0,2x)	y=62509ln(x)-169021	0,2437	19580	J(4,64)
73	y=1,2x^4	y=1,1918x^(4,0042)	1	0,0032	N(5,1)
74	y=0,8x^6	y=0,8x^6	1	1,3*10^(-6)	N(20,5)
75	y=0,9x^8	y=0,9x^8	1	4*10^(-9)	N(3,10)
76	y=1,2x^4	y=106,49x^2-505,02x+630,24	0,9949	1,091	N(4,1)
77	y=0,8x^6	y=10^6x^2-5*10^7x+6*10^8	0,1436	26,224606	N(30,5)
78	y=0,9x^8	y=9*10^11x^2-7*10^13x+10^15	0,517	381914,31	N(50,10)
79	y=1,2x^4	y=10,5078exp(0,8346x)	0,876	0,1842	N(5,1)
80	y=0,01x^6	y=9696,5exp(0,2197x)	0,9923	5,1447	N(20,5)
81	y=0,9x^8	y=9699,5exp(0,389x)	0,5853	0,9893	N(35,10)
82	y=1,2x^4	y=2786,4ln(x)-3516,8	0,7832	1,9856	N(5,1)
83	y=0,8x^6	y=3*10^8ln(x)-9*10^8	0,5015	165,25	N(20,5)
84	0,9x^8	y=-0,0256ln(x)+0,1642	8*10^-5	1,167	N(35,10)
85	y=1,2x^4	y=2,2572x^(3,986)	1	0,3787	J(0,6)
86	y=0,8x^6	y=0,8353x^(5,9888)	1	0,0103	J(5,35)
87	y=0,9x^8	y=0,9104x^(7,997)	1	0,0034	J(2,62)
88	y=1,2x^4	y=121,57x^2-558,28x+595,53	0,9921	21,942	J(1,7)
89	y=0,8x^6	y=7192x^3-2*10^6*x^2+2*10^7x-3*10^7	0,6429	8*10^6	J(0,30)
90	y=0,9*x^8	y=10^11x^2-5*10^12x+4*10^13	0,7769	3*10^8	J(3,63)
91	y=1,2x^4	y=7,4447exp(0,8652x)	0,802	0,3186	J(2,8)
92	y=0,4x^5+2	y=19301exp(0,21x)	0,9967	5465,7	J(0,30)
93	y=0,9x^8	y=862088exp(0,2927x)	0,8476	51,142	J(4,64)
94	y=1,2x^4	y=2903ln(x)-3153,9	0,7184	10,124	J(2,8)
95	y=0,8x^6	y=2*10^8ln(x)-2*10^18	0,1774	10^8	J(0,30)
96	0,9x^8	y=2*10^12ln(x)-4*10^12	0,3347	8*10^7	J(4,64)

Tab. 2. Przykłady wskazujące na zgodność lub brak zgodności wniosków dla różnych typów danych i dopasowywanych funkcji

Table 2. Examples illustrating consistency and lack of consistency of conclusions for different kinds of data and fitting functions

a) zgodne wnioski

a) consistent conclusions

wielomian – wielomian (A)		wykładnicza – logarytmiczna (B)	
R^2	*RMS*	R^2	*RMS*
0,9818	0,1679	0,1865	995,1100
0,9999	0,7940	0,1199	173,8200
1	0,8147	0,1260	5702,7000
1	0,4805	0,4124	220,0000
1	0,0024	0,0572	419942,0000
1	0,0080	0,2437	19580,0000

b) brak zgodności

b) lack of consistency

potęgowa – wykładnicza (C)		wykładnicza – wielomian (D)	
R^2	*RMS*	R^2	*RMS*
0,8760	0,1842	0,9008	21,7590
0,9923	5,1447	0,9916	0,0889
0,5853	0,9893	0,7782	138,6800
0,8020	0,3186	0,9621	75,8030
0,9967	5465,7000	0,9336	136,8200
0,8476	51,1420	0,6680	82216,0000

Tab. 3. Wartości współczynników R^2 i *RMS* przy rozkładach *x*: normalnym z dowolną wartością μ i $\sigma = 1$ oraz jednostajnym na przedziale o długości 6
Table 3. Values of coefficients of R^2 and RMS under normal distributions of x with arbitrary μ and σ^2=1 and unified distribution on interval with length 6

$N(*,1)$		$J(0,6) \vee J(1,7) \vee J(2,8)$	
R^2	*RMS*	R^2	*RMS*
0,9818	0,1679	1,0000	0,4805
0,7639	73,1690	0,9531	0,1226
0,9445	0,0686	0,9980	0,0306
0,9914	0,0438	0,8630	0,5346
0,7986	10,0100	0,4610	33,3670
0,8411	0,2281	0,3431	2,2850
0,1941	11,5380	0,1081	1,5978
0,2815	0,1914	0,9621	75,8030
1,0000	0,0023	0,4124	220,0000
0,9008	21,7590	1,0000	0,3787
0,1865	995,1100	0,8020	0,3186
0,7019	0,1649	0,7184	10,1240
1,0000	0,0032	0,6954	19,0910
0,9949	1,0910	0,4645	0,9768
0,8760	0,1842	1,0000	0,0276
0,7832	1,9856	0,9921	21,9420

Tab. 4. Współczynniki korelacji między R^2 i *RMS* przy różnych rozkładach *x* i różnych przedziałach zmienności
Table 4. Coefficients of correlations between R^2 and RMS under different distributions of x and different intervals of variation

	$N(*,1)$	$N(*,5)$	$N(*,10)$	$J(0,6) \vee J(1,7) \vee J(2,8)$	$J(0,30) \vee J(5,35)$	$J(2,62) \vee J(3,63) \vee J(4,64)$
Współczynnik korelacji	0,0763	-0,0670	0,0653	-0,0810	-0,0800	-0,0776

5. Podsumowanie i wnioski

Przeprowadzone porównania dwóch wybranych współczynników badających dobroć dopasowania nie przyniosło jednoznacznej odpowiedzi, który z nich jest lepszy, czy też który z nich powinien być wykorzystywany w konkretnych rozważanych tu przypadkach. Można jednak zasugerować następujące stwierdzenia:

1. Ze względu na ograniczony zakres wartości, jaki może przyjmować współczynnik $R^2 \in \langle 0,1 \rangle$ jest on wygodniejszy do interpretacji.
2. Wnioski o dobroci dopasowania modelu regresyjnego uzyskane przy użyciu R^2 i *RMS* mogą się zdecydowanie różnić. Nie stwierdzono jednoznacznie, kiedy występuje zgodność we wnioskowaniu, a kiedy wnioski mogą się różnić.
3. Ze wstępnych badań można przyjąć stwierdzenie, że największa zgodność wniosków uzyskanych przy użyciu obu wskaźników, co do dobroci dopasowania modelu, jest w przypadku dopasowania wielomianu do danego wielomianu i przy losowanych wartościach *x* z rozkładu normalnego o małym zakresie wartości oraz dopasowania funkcji logarytmicznej do wykładniczej bez względu na rozkład cechy *x*.
4. Zmienność wartości współczynnika *RMS* jest też w przypadkach wskazanych we wniosku 3. mniejsza niż w pozostałych.

6. Literatura

[1] Barrett J.P.: The Coefficient of Determination – Some Limitations The American Statistician, 1974, 28,1, pp 19-20.
[2] Białobrzeski J.: Wybrane problemy modelowania procesów konwekcyjnego suszenia owoców i warzyw. Rozprawa habilitacyjna. Olsztyn, 2006.
[3] Kornacki A., Wesołowska-Janczarek M.: O weryfikowaniu poprawności matematycznych modeli procesów w oparciu o dane empiryczne. Problemy Inżynierii Rolniczej, 2008, (3)61, s 5-18.
[4] Magge, L.: R^2 Measures Based on Wald and Likelihood Ratio Joint Significance Tests. The American Statistician. 1990, 44, 1, pp 250-253.
[5] Nagelkere N.J.D.: A note on a general definition of the coefficient of determination. Biometrika, 1991, 78, 3, pp. 691-692.
[6] Rencher A.C., Pun F.C.: Inflation of R^2 in Best Subset Regression. Technometrics, 1980, 22, 1, 49-53.
[7] Ryan T.P.: Modern Regression Methods. John Wiley & Sons, New York, 1997.