

Ekonometria

Regresja liniowa, dobór postaci analitycznej, transformacja
liniowa

Paweł Cibis
pawel@cibis.pl

24 marca 2007

- 1 Regresja liniowa
 - Regresja – wzory
 - Regresja – wykres
 - Regresja – funkcje
 - Regresja – REGLINP
 - Regresja – Analiza danych
 - Regresja – „Krok po kroku”
- 2 Dobór postaci analitycznej
 - Metoda aprioryczna
 - Metoda heurystyczna
 - Metoda oceny wzrokowej rozrzutu punktów
- 3 Transformacja liniowa
 - Etapy transformacji liniowej
 - Przydatne funkcje
 - Przykłady transformacji liniowej
- 4 Literatura

- 1 Regresja liniowa
 - Regresja – wzory
 - Regresja – wykres
 - Regresja – funkcje
 - Regresja – REGLINP
 - Regresja – Analiza danych
 - Regresja – „Krok po kroku”
- 2 Dobór postaci analitycznej
 - Metoda aprioryczna
 - Metoda heurystyczna
 - Metoda oceny wzrokowej rozrzutu punktów
- 3 Transformacja liniowa
 - Etapy transformacji liniowej
 - Przydatne funkcje
 - Przykłady transformacji liniowej
- 4 Literatura

Estymacja parametrów prostych regresji

$$\hat{Y} = a_1 X + b_1$$

$$a_1 = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

$$b_1 = \bar{y} - a_1 \bar{x}$$

Estymacja parametrów prostych regresji

$$\hat{Y} = a_1 X + b_1$$

$$a_1 = \frac{n \sum_{i=1}^n y_i x_i - \sum_{i=1}^n y_i \sum_{i=1}^n x_i}{n \sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2}$$

$$b_1 = \frac{1}{n} \sum_{i=1}^n y_i = \frac{a_1}{n} \sum_{i=1}^n x_i$$

Estymacja parametrów prostych regresji

$$\hat{X} = a_2 Y + b_2$$

$$a_2 = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (y_i - \bar{y})^2}$$

$$b_2 = \bar{x} - a_2 \bar{y}$$

Estymacja parametrów prostych regresji

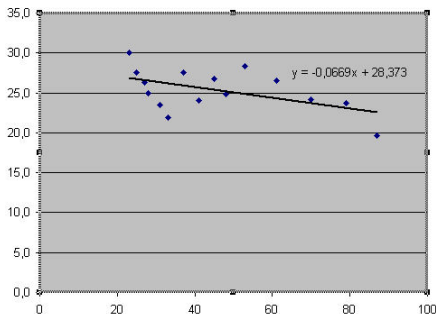
$$\hat{X} = a_2 Y + b_2$$

$$a_2 = \frac{n \sum_{i=1}^n y_i x_i - \sum_{i=1}^n y_i \sum_{i=1}^n x_i}{n \sum_{i=1}^n y_i^2 - (\sum_{i=1}^n y_i)^2}$$

$$b_2 = \frac{1}{n} \sum_{i=1}^n x_i = \frac{a_2}{n} \sum_{i=1}^n y_i$$

Estymacja za pomocą wykresu

- 1 Utwórz wykres korelacyjny – XY (Punktowy)
- 2 Wykres/Dodaj linię trendu...
 - Typ – Liniowy
 - Opcje – Wyświetl równanie na wykresie



Estymacja za pomocą funkcji Excela

$a = \text{NACHYLENIE}(\text{zakres_obja\u015bnianej}; \text{zakres_obja\u015bniaj\u0105cej})$

$b = \text{ODCI\u0118TA}(\text{zakres_obja\u015bnianej}; \text{zakres_obja\u015bniaj\u0105cej})$

Estymacja za pomocą formuły tablicowej REGLINP

REGLINP(zakres_objaśnianej; zakres_objaśniającej; wyraz_wolny; statystyki)

- Wpisz formułę REGLINP wraz z paramterami w puste pole arkusza i naciśnij ENTER
 - zamiast jednej zmiennej objaśniającej, może być zestaw wielu zmiennych objaśniających
 - w polu "wyraz_wolny" (w Excelu "Stała") wpisz 1
 - w polu "statystyki" ("Statystyka") wpisz 1, jeśli chcesz otrzymać statystyki modelu

Estymacja za pomocą formuły tablicowej REGLINP

- Zaznacz obszar począwszy od pola z formułą REGLINP o szerokości równej liczbie zmiennych objaśniających + 1 i wysokości równej 5 (jeśli argument "statystyki" został pominięty wysokość jest równa 1)
- Naciśnij klawisz F2 (musi być zaznaczony wspomniany powyżej obszar, a aktywna komórka powinna zawierać formułę REGLINP)
- Naciśnij kombinację klawiszy: SHIFT+CTRL+ENTER

Estymacja za pomocą formuły tablicowej REGLINP

- I wiersz: oszacowania parametrów kolejnych zmiennych (wyraz wolny w ostatniej kolumnie)
- II wiersz: standardowe błędy szacunku powyższych parametrów
- III wiersz: współczynnik determinacji R^2 oraz standardowy błąd oceny zmiennej objaśnianej
- IV wiersz: statystyka F i stopnie swobody (istotność parametrów strukturalnych)
- V wiersz: regresyjna suma kwadratów i resztkowa suma kwadratów

Estymacja za pomocą formuły tablicowej REGLINP

E1	fx {=REGLINP(C2:C16;B2:B16;1;1)}					
	A	B	C	D	E	F
1	Lata	X_t	Y_t		-0,06685	28,37294
2	1987	87	19,6		0,03111	1,552085
3	1988	79	23,7		0,26211	2,365109
4	1989	70	24,2		4,61779	13
5	1990	61	26,5		25,83072	72,71862
6	1991	53	28,3			
7	1992	48	24,8			
8	1993	45	26,8			
9	1994	41	24,0			
10	1995	37	27,5			
11	1996	33	21,9			
12	1997	31	23,5			
13	1998	28	25,0			
14	1999	27	26,3			
15	2000	25	27,5			
16	2001	23	30,0			

Estymacja za pomocą pakietu Analiza Danych

- 1 Jeżeli w menu Narzędzia nie ma opcji "Analiza Danych...", doinstaluj ją (Narzędzia/Dodatki...)
- 2 Narzędzia/Analiza Danych.../Regresja
 - zaznacz odpowiednie zakresy wejściowe dla danych oraz wybierz statystyki do podsumowania
 - wybierz opcję wyścia (w przypadku wyboru pola "Zakres" zaznaczonym polem będzie lewy górny róg podsumowania – pamiętaj, by poniżej i na prawo od tego pola było odpowiednio dużo wolnego miejsca)

Estymacja za pomocą pakietu Analiza Danych

	A	B	C	D	E	F	G	H	I
1	PODSUMOWANIE - WYJŚCIE								
2									
3	<i>Statystyki regresji</i>								
4	Wielokrotn	0,511966							
5	R kwadrat	0,26211							
6	Dopasowa	0,205349							
7	Błąd stand	2,365109							
8	Obserwacj	15							
9									
10	ANALIZA WARIANCJI								
11		<i>df</i>	<i>SS</i>	<i>MS</i>	<i>F</i>	<i>Istotność F</i>			
12	Regresja	1	25,83072	25,83072	4,61779	0,051058			
13	Resztkowy	13	72,71862	5,59374					
14	Razem	14	98,54933						
15									
16		<i>Współczynnik standardowy</i>	<i>t Stat</i>	<i>Wartość-p</i>		<i>Dolne 95%</i>	<i>Górne 95%</i>	<i>Dolne 95,0%</i>	<i>Górne 95,0%</i>
17	Przecięcie	28,37294	1,552085	18,28054	1,18E-10	25,01987	31,72602	25,01987	31,72602
18	X3t	-0,06685	0,03111	-2,1489	0,051058	-0,13406	0,000357	-0,13406	0,000357
19									

Szacowanie parametrów modelu „krok po kroku”

- Metoda ta polega na obliczaniu kolejnych elementów składowych podanych na początku prezentacji wzorów – tak samo, jak przy liczeniu „ręcznym”.

Szacowanie parametrów modelu „krok po kroku”

- Metoda ta polega na obliczaniu kolejnych elementów składowych podanych na początku prezentacji wzorów – tak samo, jak przy liczeniu „ręcznym”.
- Zaletą jest możliwość automatycznego przeliczenia wyników w przypadku korekty danych wejściowych, co jest niemożliwe w pakiecie Analiza Danych oraz możliwość łatwego wglądu do wyników cząstkowych.

Szacowanie parametrów modelu „krok po kroku”

- Metoda ta polega na obliczaniu kolejnych elementów składowych podanych na początku prezentacji wzorów – tak samo, jak przy liczeniu „ręcznym”.
- Zaletą jest możliwość automatycznego przeliczenia wyników w przypadku korekty danych wejściowych, co jest niemożliwe w pakiecie Analiza Danych oraz możliwość łatwego wglądu do wyników cząstkowych.
- Wadą jest duża pracochłonność i podatność na błędy w porównaniu do poprzednich metod oraz konieczność dobrego rozplanowania arkusza (szczególnie w przypadku dużej próby).

Szacowanie parametrów modelu „krok po kroku”

	A	B	C	D	E	F	G	H	I	J	K
73											
74	X_{3t}	Y_t	$X_{3t} - X_{3t\bar{s}}$	$Y_t - Y_{t\bar{s}}$	Licznik	$(X_{3t} - X_{3t\bar{s}})^2$		X_{3t}	Y_t	$X_{3t} * Y_t$	X_{3t}^2
75	87	19,6	41,13	-5,71	-234,73	1691,95		87	19,6	1705,20	7569,00
76	79	23,7	33,13	-1,61	-53,23	1097,82		79	23,7	1872,30	6241,00
77	70	24,2	24,13	-1,11	-26,71	582,42		70	24,2	1694,00	4900,00
78	61	26,5	15,13	1,19	18,06	229,02		61	26,5	1616,50	3721,00
79	53	28,3	7,13	2,99	21,35	50,88		53	28,3	1499,90	2809,00
80	48	24,8	2,13	-0,51	-1,08	4,55		48	24,8	1190,40	2304,00
81	45	26,8	-0,87	1,49	-1,29	0,75		45	26,8	1206,00	2025,00
82	41	24,0	-4,87	-1,31	6,36	23,68		41	24,0	984,00	1681,00
83	37	27,5	-8,87	2,19	-19,45	78,62		37	27,5	1017,50	1369,00
84	33	21,9	-12,87	-3,41	43,83	165,55		33	21,9	722,70	1089,00
85	31	23,5	-14,87	-1,81	26,86	221,02		31	23,5	728,50	961,00
86	28	25,0	-17,87	-0,31	5,48	319,22		28	25,0	700,00	784,00
87	27	26,3	-18,87	0,99	-18,74	355,95		27	26,3	710,10	729,00
88	25	27,5	-20,87	2,19	-45,77	435,42		25	27,5	687,50	625,00
89	23	30,0	-22,87	4,69	-107,32	522,88		23	30,0	690,00	529,00
90					-386,39	5779,73		688	379,6	17024,6	37336,00
91	średnia X_{3t}	45,87		a_1	-0,07			n	15,0	a_1	-0,07
92	średnia Y_t	25,31		b_1	28,37					b_1	28,37
93											

- 1 Regresja liniowa
 - Regresja – wzory
 - Regresja – wykres
 - Regresja – funkcje
 - Regresja – REGLINP
 - Regresja – Analiza danych
 - Regresja – „Krok po kroku”
- 2 Dobór postaci analitycznej
 - Metoda aprioryczna
 - Metoda heurystyczna
 - Metoda oceny wzrokowej rozrzutu punktów
- 3 Transformacja liniowa
 - Etapy transformacji liniowej
 - Przydatne funkcje
 - Przykłady transformacji liniowej
- 4 Literatura

Metoda aprioryczna

Polega na doborze postaci analitycznej modelu na podstawie informacji pozastatystycznych, doyczących związku łączącego zmienną objaśnianą ze zmiennymi objaśniającymi. Źródłami tych informacji mogą być:

- teoria ekonomii i ekonomik branżowych,
- opinie ekspertów,
- tradycje i doświadczenia badawcze.

Metoda aprioryczna

Wnioski wynikające z apriorycznej wiedzy o zależnościach między zmiennymi:

- stałe absolutne przyrosty zmiennej objaśnianej – funkcja liniowa,
- stałe względne przyrosty zmiennej objaśnianej – funkcja wykładnicza,
- stałe współczynniki elastyczności – funkcja potęgowa,
- coraz większe przyrosty zmiennej objaśnianej przy wzroście zmiennej objaśniającej o każdą kolejną jednostkę – funkcja potęgowa, wykładnicza lub kwadratowa,
- coraz mniejsze przyrosty zmiennej objaśnianej przy wzroście zmiennej objaśniającej o każdą kolejną jednostkę – funkcja potęgowa lub logarytmiczna.

Metoda heurystyczna

Polega na zastosowaniu modeli o różnych postaciach analitycznych i wyborze jednego z nich na podstawie wyróżnionego kryterium dobroci dopasowania modelu do rzeczywistości. Istnieją 2 warianty tej metody:

- metoda kolejnych przybliżeń,
- metoda zadowalającego wyboru.

Metoda heurystyczna – metoda kolejnych przybliżeń

Analizie poddawanych jest wiele funkcji matematycznych, na podstawie których budowane są modele ekonometryczne. W etapie weryfikacji oblicza się dla każdego modelu współczynnik determinacji R^2 (przyjmujący wartości z przedziału $[0,1]$). Do modelu dobierana jest funkcja o maksymalnej wartości współczynnika determinacji R^2 .

Metoda heurystyczna – metoda zadowalającego wyboru

Należy ustalić wartość krytyczną współczynnika determinacji (R_*^2). Analizę rozpoczyna się od najprostszych funkcji, porównując wartość R^2 badanego modelu z wartością krytyczną R_*^2 i jeżeli ta pierwsza jest jej co najmniej równa, należy przyjąć daną postać analityczną modelu.

Metoda heurystyczna – warunki stosowalności

- brak informacji pozastatystycznych o związku łączącym zmienne modelu,
- występowanie wielu zmiennych objaśniających,
- jeśli metoda aprioryczna lub metoda oceny wzrokowej nie dają jednoznacznego rozstrzygnięcia, co do postaci analitycznej modelu – metoda heurystyczna stosowana jest wtedy pomocniczo.

Wady: subiektywizm i pracochłonność

Metoda oceny wzrokowej rozrzutu punktów

Polega na przedstawieniu na wykresie korelacyjnym rozrzutu punktów empirycznych i przypisaniu badanej zależności funkcji, której przebieg zmienności jest najbardziej zbliżony do uzyskanej smugi punktów.

Warunkiem stosowania tej metody jest występowanie tylko jednej zmiennej objaśniającej.

Metoda aproksymacji segmentowej

Jest to szczególny przypadek metody oceny wzrokowej. Ma zastosowanie, gdy rozrzut punktów jest nieciągły lub funkcja jest powikłana.

Wykres rozrzutu należy podzielić na części zwane segmentami. Każdemu segmentowi przypisuje się oddzielną funkcję opisującą związek między zmiennymi, zwaną aproksymantą segmentową. Wartość zmiennej objaśniającej, dla której dokonuje się cięcia wykresu nazywamy modulatorem.

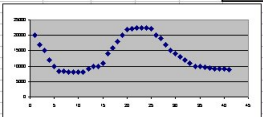
Metoda aproksymacji segmentowej

Funkcja obliczająca wartość oceny zmiennej objaśnianej w danym punkcie wynikającą z linii regresji oszacowanej na podstawie zadanego ciągu obserwacji:

$$= REGLINW(Znane_Y; Znane_X; Nowe_X; Wyraz_wolny)$$

Metoda aproksymacji segmentowej

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	t	Y	t	t ²	REGLINP I segment				ysr	y ^r	(y ^r -ysr) ²	(y-ysr) ²	(y-y ^r) ²	
2	1	20000	1	1	152,88464	-2979,24	22071,88			1,3682,46	19245,32	30945393,5	39911268,5	569539,2
3	2	17000	2	4	5,6948247	117,236	509,1696			R ²	16724,73	9255409,31	11006049	75771,46
4	3	15000	3	9	0,9785147	663,2943	#N/DI			0,982962	14509,92	884676,359	1735902,85	240183,2
5	4	12000	4	16	364,34717	16	#N/DI			Phi ²	12600,87	1169853,57	2830683,14	361039,6
6	5	10000	5	25	320595872	7039349	#N/DI			0,017038	10997,59	7208568,98	13560536,8	995177
7	6	8417	6	36						suma	9700,075	15659419,7	27725105	1646281
8	7	8267	7	49							8708,333	24741972,1	29327244	194774,9
9	8	8117	8	64							8022,361	32036761,5	30974383	8956,576
10	9	8100	9	81							7642,158	36485292,2	31163997,8	209619,5
11	10	8110	10	100							7567,724	37390037,9	31052348,5	294063,2
12	11	8120	11	121							7799,06	34614441,3	30940999,2	103002,8
13	12	9000	12	144							8336,164	28582914	21925463,6	440677,8
14	13	9800	13	169							9179,038	20208037	15073522,2	385933,3
15	14	10000	14	196							10327,58	11254580,1	13560536,8	107375,3
16	15	11000	15	225							11782,09	3611402,57	7195609,97	611871,6
17	16	14000	16	256							13542,28	198574413	100829,483	209511
18	17	18000	17	289							15806,23	3708587,01	5370975,82	153485,8
19	18	18000	18	324							17979,85	19468372,6	1864112,2	402,0679
20	19	20000	19	361							20657,44	48650264,8	39911268,5	43224,2
21	20	22000	20	400	8000	5,8676593	-518,705	14022,98	-97385,7		22532,96	76331368,2	69181414,8	284051,1
22	21	22000	21	441	5261	0,8374662	76,76777	2294,236	22314,23		22688,14	81102201,2	72549429,5	236280,1
23	22	22500	22	484	10648	0,8830374	779,0056	#N/DI	#N/DI		22545,23	78548606,2	77748951,4	2045,634
24	23	22400	23	529	12167	347,71875	18	#N/DI	#N/DI		22139,44	71520419,5	75985444,1	67892,54
25	24	22400	24	576	13824	633039091	10923295	#N/DI	#N/DI		21505,97	61207309	75985444,1	799263,5
26	25	22200	25	625	15825						20860,04	48966090,1	72548429,5	2010276
27	26	20000	26	676	17576						19696,95	36172801,1	39911268,5	91902,19
28	27	19000	27	729	19683						18591,6	24099578,7	28276195,3	186794,2
29	28	17000	28	784	21952						17399,49	13816322,2	11006049	159596,1
30	29	15000	29	841	24389						16155,75	6117145,54	1735902,85	1335758
31	30	14000	30	900	27000						14895,57	1471619,3	100829,483	802039,5
32	31	13000	31	961	29791						13654,15	801,587772	485756,312	427913,6
33	32	12000	32	1024	32768						12466,71	1478058,16	2830683,14	217817,5
34	33	11000	33	1089	35937						11368,45	5354671,71	7195609,97	135753,2
35	34	10000	34	1156	39304						10394,57	10810240,4	13560536,8	155685,8
36	35	10000	35	1225	42875						9580,285	16827865,3	13560536,8	176160,4
37	36	9600	36	1296	46566						8960,798	22294127,6	16666507,5	408579,6
38	37	9500	37	1369	50653						8571,313	26123854,3	17493000,2	862458,8
39	38	9200	38	1444	54872						8447,039	27409673,6	20092478,3	566951
40	39	9200	39	1521	59319						8623,179	25596358,8	20092478,3	332722,5
41	40	9000	40	1600	64000						9134,941	20679962,7	21925463,6	18209
42	41	8850	41	1681	68921						10017,53	13431739,5	23352702,7	1363129
43												1036329214	1054291858	17962644



- 1 Regresja liniowa
 - Regresja – wzory
 - Regresja – wykres
 - Regresja – funkcje
 - Regresja – REGLINP
 - Regresja – Analiza danych
 - Regresja – „Krok po kroku”
- 2 Dobór postaci analitycznej
 - Metoda aprioryczna
 - Metoda heurystyczna
 - Metoda oceny wzrokowej rozrzutu punktów
- 3 Transformacja liniowa
 - Etapy transformacji liniowej
 - Przydatne funkcje
 - Przykłady transformacji liniowej
- 4 Literatura

Etapy transformacji liniowej

- 1 Dobór postaci analitycznej modelu (np. analiza wykresu korelacyjnego)
- 2 Transformacja liniowa funkcji i dokonanie podstawień zmiennych i/lub parametrów
- 3 Szacowanie parametrów modelu liniowego
- 4 Obliczenie parametrów modelu nieliniowego

Przydatne funkcje

$$\ln(x) = \text{LN}(x)$$

$$\log_a x = \text{LOG}(x; [a])$$

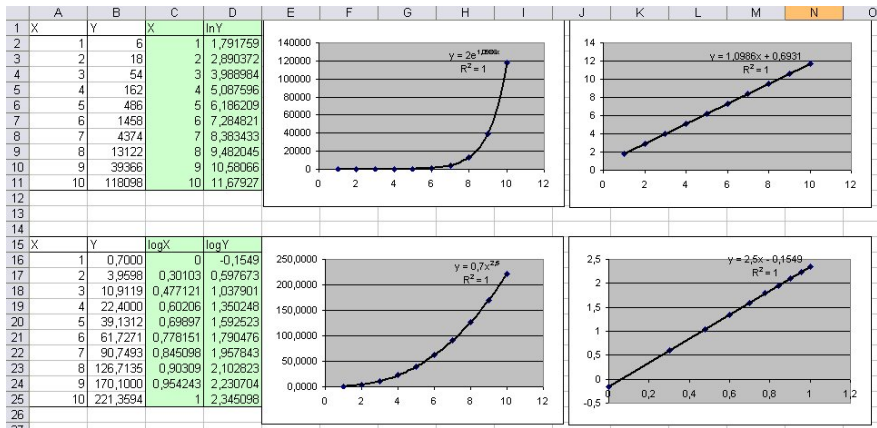
$$\log x = \text{LOG}(x)$$

$$e^x = \text{EXP}(x)$$

$$\sin(x) = \text{SIN}(x)$$

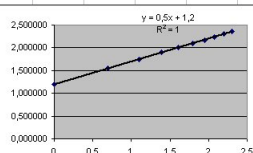
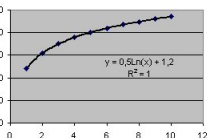
$$\cos(x) = \text{COS}(x)$$

Przykłady transformacji liniowej 1

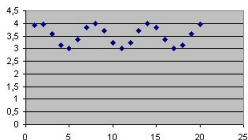


Przykłady transformacji liniowej 2

28				
29	X	Y	lnX	Y
30	1	1,200000	0	1,200000
31	2	1,546574	0,693147	1,546574
32	3	1,749306	1,098612	1,749306
33	4	1,893147	1,386294	1,893147
34	5	2,004719	1,609438	2,004719
35	6	2,095880	1,791759	2,095880
36	7	2,172955	1,94591	2,172955
37	8	2,239721	2,079442	2,239721
38	9	2,298612	2,197225	2,298612
39	10	2,351293	2,302585	2,351293



43	X	Y	sinx	Y	cosx	Y
44	1	3,920735	0,841471	3,920735	0,540302	3,920735
45	2	3,954649	0,909297	3,954649	-0,41615	3,954649
46	3	3,57056	0,14112	3,57056	-0,98999	3,57056
47	4	3,121599	-0,7568	3,121599	-0,65364	3,121599
48	5	3,020538	-0,95892	3,020538	0,283662	3,020538
49	6	3,360292	-0,27942	3,360292	0,96017	3,360292
50	7	3,828493	0,656987	3,828493	0,753902	3,828493
51	8	3,994679	0,989358	3,994679	-0,1455	3,994679
52	9	3,706059	0,412118	3,706059	-0,91113	3,706059
53	10	3,227989	-0,54402	3,227989	-0,83907	3,227989
54	11	3,000005	-0,99999	3,000005	0,004426	3,000005
55	12	3,231714	-0,53657	3,231714	0,843854	3,231714
56	13	3,710084	0,420167	3,710084	0,907447	3,710084
57	14	3,995304	0,990607	3,995304	0,136737	3,995304
58	15	3,825144	0,650288	3,825144	-0,75969	3,825144
59	16	3,356048	-0,2879	3,356048	-0,95766	3,356048
60	17	3,019301	-0,9614	3,019301	-0,27516	3,019301
61	18	3,124506	-0,75099	3,124506	0,660317	3,124506
62	19	3,574939	0,149877	3,574939	0,988705	3,574939
63	20	3,956473	0,912945	3,956473	0,408082	3,956473



- 1 Regresja liniowa
 - Regresja – wzory
 - Regresja – wykres
 - Regresja – funkcje
 - Regresja – REGLINP
 - Regresja – Analiza danych
 - Regresja – „Krok po kroku”
- 2 Dobór postaci analitycznej
 - Metoda aprioryczna
 - Metoda heurystyczna
 - Metoda oceny wzrokowej rozrzutu punktów
- 3 Transformacja liniowa
 - Etapy transformacji liniowej
 - Przydatne funkcje
 - Przykłady transformacji liniowej
- 4 Literatura

Literatura



Strahl D., Sobczak E., Markowska M., Bal-Domańska B. *Modelowanie ekonometryczne z Excelem*. Wrocław: AE 2004.



Ekonometria. Metody, przykłady, zadania. Red. J. Dziechciarz. Wrocław: AE 2002.



Ekonometria. Metody i analiza problemów ekonomicznych. Red. K. Jajuga. Wrocław: AE 1999.

