



UNIwersytet
EKONOMICZNY
W POZNANIU

Załącznik nr 1
do Zarządzenia nr 92/2021 Rektora UEP
z dnia 28 października 2021 roku

Tomasz Firlej

Liczba samochodów osobowych zarejestrowanych
na 1000 ludności w latach 2005-2022

Projekt ekonometryczny

Kierunek: Informatyka i ekonometria



UNIwersytet
EKONOMICZNY
W POZNANIU

Poznań 2024

Zmienna zależna:

Y - Liczba samochodów osobowych zarejestrowanych na 1000 ludności

Źródło: gus.pl

Zmienne objaśniające:

X₁ - Drogi ekspresowe i autostrady na 1000 ludności

Drogi ekspresowe i autostrady są kluczowe dla mobilności i dostępności transportowej. Większa liczba takich dróg może zachęcać do posiadania samochodu ze względu na lepsze warunki podróżowania i skrócenie czasu dojazdów.

Źródło: gus.pl

X₂ - Linie kolejowe ogółem na 1000 mieszkańców

Transport kolejowy może stanowić alternatywę dla transportu samochodowego. Istnienie rozbudowanej sieci kolejowej może zmniejszać potrzebę posiadania samochodu, dlatego ta zmienna może mieć wpływ na liczbę samochodów osobowych.

Źródło: gus.pl

X₃ - Cena przeciętnego mycia mechanicznego nadwozia samochodu osobowego

Cena mycia samochodu może być wskaźnikiem ogólnych kosztów utrzymania samochodu. Wysokie koszty mogą zniechęcać do posiadania samochodu, podczas gdy niższe koszty mogą działać odwrotnie.

Źródło: gus.pl

X₄ - Przeciętny miesięczny dochód rozporządzalny na 1 osobę

Dochód rozporządzalny jest jednym z najważniejszych czynników wpływających na zdolność i skłonność do zakupu samochodu. Wyższy dochód pozwala na zakup i utrzymanie samochodu, dlatego jest to kluczowa zmienna w modelu.

Źródło: gus.pl

X₅ - Stopa bezrobocia rejestrowanego

Stopa bezrobocia odzwierciedla ogólną sytuację ekonomiczną i może wpływać na zdolność finansową do zakupu samochodu. Wysoka stopa bezrobocia może ograniczać liczbę osób, które mogą sobie pozwolić na posiadanie samochodu.

Źródło: gus.pl

X₆ - Małżeństwa na 1000 mieszkańców

Liczba małżeństw może być wskaźnikiem stabilności społecznej i ekonomicznej. Małżeństwa często prowadzą do wzrostu liczby samochodów w gospodarstwie domowym ze względu na większe potrzeby transportowe.

Źródło: statista.com

X_7 - Liczba kradzieży samochodów

Liczba kradzieży samochodów może wpływać na decyzję o posiadaniu samochodu. Wysoka liczba kradzieży może zniechęcać do zakupu samochodu z obawy przed utratą.

Źródło: statista.com

Model hipotetyczny

$$Y_t = \beta_0 + \beta_1 X_{t,1} + \beta_2 X_{t,2} + \beta_3 X_{t,3} + \beta_4 X_{t,4} + \beta_5 X_{t,5} + \beta_6 X_{t,6} + \beta_7 X_{t,7} + \varepsilon_t$$

Materiał statystyczny:

	Y	X1	X2	X3	X4	X5	X6	X7
2005	323,4	0,21	5,3	13,85	731,84	17,6	5,40	45 292
2006	351,1	0,25	5,3	13,99	802,43	14,8	5,90	30 529
2007	382,7	0,26	5,3	13,98	894,51	11,2	6,50	21 284
2008	421,6	0,32	5,3	14,45	1 006,57	9,5	6,80	17 669
2009	432,2	0,36	5,3	14,90	1 071,67	12,1	6,60	17 271
2010	447,4	0,40	5,2	15,18	1 147,18	12,4	5,90	16 539
2011	470,3	0,47	5,3	15,63	1 183,66	12,5	5,40	16 575
2012	486,4	0,63	5,2	16,34	1 232,85	13,4	5,30	16 230
2013	503,7	0,71	5,0	16,74	1 254,89	13,4	4,70	14 800
2014	519,9	0,78	5,0	17,04	1 293,32	11,4	4,90	13 791
2015	539,1	0,79	5,0	17,36	1 337,87	9,7	4,90	12 036
2016	564	0,83	5,0	17,69	1 425,75	8,2	5	11 448
2017	585,5	0,89	5,0	17,98	1 548,87	6,6	5	10 047
2018	610	0,97	5,0	18,34	1 643,14	5,8	5	8 745
2019	634,7	1,07	5,1	19,19	1 768,65	5,2	4,80	8 672
2020	659,4	1,12	5,1	19,95	1 874,24	6,3	3,80	8 784
2021	682,4	1,20	5,1	20,81	2 014,61	5,8	4,40	8 383
2022	700,6	1,29	5,1	23,25	2 194,62	5,2	4,10	7 377

Źródła: gus.pl, statista.com

Weryfikacja statystyczna modelu:

Regresja krokowa:

Model 1:

	Współczynniki i błęd standardowy		t Stat	Wartość-p	Dołne 95%	Górne 95%	Dołne 95,0%	Górne 95,0%
Przecięcie	589,706654	164,044651	3,59479355	0,00489	224,192393	955,220914	224,192393	955,220914
X1	125,351512	46,2037449	2,71301629	0,02182034	22,4031534	228,299872	22,4031534	228,299872
X2	-22,837445	32,4631561	-0,7034881	0,4978037	-95,169864	49,4949748	-95,169864	49,4949748
X3	-11,153268	4,63614188	-2,4057218	0,03695257	-21,483236	-0,8232998	-21,483236	-0,8232998
X4	0,17523221	0,04246058	4,12693892	0,0020542	0,08062415	0,26984027	0,08062415	0,26984027
X5	-1,9590219	1,37256946	-1,4272661	0,1839748	-5,0172972	1,09925346	-5,0172972	1,09925346
X6	-8,915726	6,74577071	-1,3216764	0,21571336	-23,94624	6,11478779	-23,94624	6,11478779
X7	-0,0014792	0,00042701	-3,4641588	0,00608027	-0,0024307	-0,0005278	-0,0024307	-0,0005278

Komentarz: zmienna X2 nie jest istotna i ma największe p-wartość, dlatego usuwamy ją z modelu

Model 2:

	Współczynniki i błęd standardowy		t Stat	Wartość-p	Dołne 95%	Górne 95%	Dołne 95,0%	Górne 95,0%
Przecięcie	484,306412	65,2542876	7,4218328	1,3231E-05	340,682693	627,93013	340,682693	627,93013
X1	145,844141	35,0309253	4,16329686	0,00158052	68,7415947	222,946688	68,7415947	222,946688
X3	-10,649371	4,4740745	-2,3802399	0,03648563	-20,496742	-0,8019989	-20,496742	-0,8019989
X4	0,15481553	0,03027391	5,11382671	0,0003367	0,0881831	0,22144795	0,0881831	0,22144795
X5	-2,0826234	1,32965651	-1,5662868	0,14557796	-5,0091777	0,84393082	-5,0091777	0,84393082
X6	-9,6074103	6,5187073	-1,4738214	0,16856267	-23,954988	4,74016772	-23,954988	4,74016772
X7	-0,0016294	0,00036124	-4,5104559	0,00088584	-0,0024245	-0,0008343	-0,0024245	-0,0008343

Komentarz: : zmienna X6 nie jest istotna i ma największe p-wartość, dlatego usuwamy ją z modelu

Model 3:

	Współczynniki i błęd standardowy		t Stat	Wartość-p	Dołne 95%	Górne 95%	Dołne 95,0%	Górne 95,0%
Przecięcie	403,668664	37,2570828	10,8346825	1,4962E-07	322,492454	484,844874	322,492454	484,844874
X1	187,410609	21,7696631	8,60879693	1,7608E-06	139,978588	234,842631	139,978588	234,842631
X3	-11,698439	4,62779295	-2,5278656	0,02652321	-21,781533	-1,6153441	-21,781533	-1,6153441
X4	0,15658421	0,0316931	4,94064035	0,00034172	0,08753088	0,22563755	0,08753088	0,22563755
X5	-0,7872811	1,04535255	-0,753125	0,46589418	-3,0649087	1,49034645	-3,0649087	1,49034645
X7	-0,0013953	0,00033995	-4,1045507	0,00146043	-0,002136	-0,0006547	-0,002136	-0,0006547

Komentarz: : zmienna X5 nie jest istotna i ma największe p-wartość, dlatego usuwamy ją z modelu

Model 4:

	Współczynniki i błęd standardowy		t Stat	Wartość-p	Dołne 95%	Górne 95%	Dołne 95,0%	Górne 95,0%
Przecięcie	401,36022	36,5074625	10,9939227	5,9294E-08	322,490643	480,229798	322,490643	480,229798
X1	188,142503	21,3828746	8,79874696	7,7725E-07	141,947611	234,337395	141,947611	234,337395
X3	-12,948519	4,24735887	-3,0486049	0,00932498	-22,12438	-3,7726579	-22,12438	-3,7726579
X4	0,16860082	0,02692405	6,26209113	2,9151E-05	0,11043496	0,22676669	0,11043496	0,22676669
X7	-0,0014667	0,000321	-4,5692653	0,00052633	-0,0021602	-0,0007732	-0,0021602	-0,0007732

Komentarz: Wszystkie zmienne są istotne

Model teoretyczny:

$$\hat{Y} = 188,14x_1 - 12,95x_3 + 0,17x_4 - 0,001x_7 + 401,36$$

$R^2 = 99,77\%$

$V = 1,19\%$

Wszystkie zmienne są istotne

Testy:

Liniowość:

Pomocnicze równanie regresji dla testu specyfikacji RESET
Estymacja KMNK, wykorzystane obserwacje 2005-2022 (N = 18)
Zmienna zależna (Y): Y

	współczynnik	błąd standardowy	t-Studenta	wartość p	
const	-688.974	488.536	-1.410	0.1861	
X1	-648.802	400.063	-1.622	0.1331	
X3	47.0756	29.3352	1.605	0.1369	
X4	-0.469121	0.339049	-1.384	0.1939	
X7	0.00403911	0.00234276	1.724	0.1126	
yhat^2	0.00823332	0.00382323	2.153	0.0543	*
yhat^3	-5.55967e-06	2.43086e-06	-2.287	0.0430	**

Hipoteza zerowa: specyfikacja poprawna

Statystyka testu: $F = 4.137671$,
z wartością $p = P(F(2,11) > 4.13767) = 0.0457$

Interpretacja:

$p = 0,04572443 < 0,05$ - model ma inną postać niż liniowy

Heteroskedastyczność:

Test White'a na heteroskedastyczność reszt (zmiennność wariancji resztowej)
Estymacja KMNK, wykorzystane obserwacje 2005-2022 (N = 18)
Zmienna zależna (Y): uhat^2

	współczynnik	błąd standardowy	t-Studenta	wartość p	
const	78951.1	37336.7	2.115	0.1248	
X1	-52330.6	23325.4	-2.243	0.1106	
X3	-1014.06	3035.05	-0.3341	0.7603	
X4	-20.0637	20.5228	-0.9776	0.4004	
X7	-5.98951	2.26571	-2.644	0.0774	*
sq_X1	-39647.0	15137.5	-2.619	0.0791	*
X2_X3	10361.5	3899.43	2.657	0.0765	*
X2_X4	-15.5438	8.10711	-1.917	0.1510	
X2_X5	-3.23019	1.16143	-2.781	0.0689	*
sq_X3	-605.979	225.819	-2.683	0.0748	*
X3_X4	5.18834	2.55171	2.033	0.1349	
X3_X5	0.538912	0.204800	2.631	0.0782	*
sq_X4	-0.0192198	0.00743966	-2.583	0.0815	*
X4_X5	-0.000391617	0.000269141	-1.455	0.2417	
sq_X7	-6.98981e-06	2.59739e-06	-2.691	0.0743	*

Wsp. determ. R-kwadrat = 0.894556

Statystyka testu: $TR^2 = 16.102005$,
z wartością $p = P(\text{Chi-kwadrat}(14) > 16.102005) = 0.307185$

Interpretacja:

$p = 0,307185 > 0,05$, Reszty są homoskedastyczne

Autokorelacja:

Stat. Durбина-Watsona = 1.779

H1: dodatnia autokorelacja składnika losowego
wartość $p = 0.0585435$

H1: ujemna autokorelacja składnika losowego
wartość $p = 0.941457$

Interpretacja: brak autokorelacji składnika losowego, jednak jest pewne zagrożenie autokorelacją dodatnią, ponieważ statystyka testowa jest bliska poziomowi krytycznego.

Rozkład normalny:

Rozkład częstości dla residual, obserwacje 1-18

liczba przedziałów = 7, średnia = $-5.68434e-14$, odch.std. = 6.15504

Przedziały	średnia	liczba	częstość	skumulowana	
< -5.1083	-6.6590	4	22.22%	22.22%	*****
-5.1083 - -2.0070	-3.5576	3	16.67%	38.89%	*****
-2.0070 - 1.0944	-0.45629	4	22.22%	61.11%	*****
1.0944 - 4.1958	2.6451	3	16.67%	77.78%	*****
4.1958 - 7.2971	5.7464	2	11.11%	88.89%	***
7.2971 - 10.398	8.8478	1	5.56%	94.44%	*
>= 10.398	11.949	1	5.56%	100.00%	*

Hipoteza zerowa: dystrybuenta empiryczna posiada rozkład normalny.

Test Doornika-Hansena (1994) - transformowana skośność i kurtoza:

Chi-kwadrat(2) = 1.796 z wartością $p = 0.40736$

Interpretacja:

Reszty mają rozkład normalny ($p = 0.407$)

Weryfikacja merytoryczna modelu:

Model regresji ma bardzo wysoką wartość $R^2 = 99,77\%$. Wskaźnik zmienności V wynosi 1.19% i wszystkie zmienne są istotne, co sugeruje że model jest dobry.

X_1 - Drogi ekspresowe i autostrady na 1000 ludności

Interpretacja:

Wzrost liczby dróg ekspresowych i autostrad na 1000 ludności o jednostkę zwiększa liczbę zarejestrowanych samochodów osobowych o 188,14 na 1000 ludności.

Oceny parametru X_1 odchylają się przeciętnie o 21,3829 od jego nieznanej rzeczywistej wartości. Przedział o granicach (141,9476, 234,3374) z prawdopodobieństwem 95% zawiera nieznaną rzeczywistą wartość parametru 2.

Merytoryczna ocena:

Wartość współczynnika jest zgodna z intuicją, ponieważ lepsza infrastruktura drogowa sprzyja posiadaniu samochodów.

X_3 - Cena przeciętnego mycia mechanicznego nadwozia samochodu osobowego

Interpretacja:

Wzrost ceny mycia samochodu o jednostkę zmniejsza liczbę zarejestrowanych samochodów o 12,95 na 1000 ludności.

Oceny parametru X_3 odchylają się przeciętnie o 4,2474 od jego nieznanej rzeczywistej wartości. Przedział o granicach (-22,1244, -3,7727) z prawdopodobieństwem 95% zawiera nieznaną rzeczywistą wartość parametru 3.

Merytoryczna ocena:

Chociaż cena mycia nadwozia jest jednym z wielu kosztów utrzymania samochodu, wpływ na liczbę samochodów wydaje się rozsądny, zwłaszcza w kontekście, gdzie ogólne koszty utrzymania mogą wpływać na decyzje konsumenckie.

X_4 - Przeciętny miesięczny dochód rozporządzalny na 1 osobę

Interpretacja:

Wzrost przeciętnego miesięcznego dochodu o jednostkę zwiększa liczbę zarejestrowanych samochodów o 0,17 na 1000 ludności.

Oceny parametru X_4 odchylają się przeciętnie o 0,0269 od jego nieznanej rzeczywistej wartości. Przedział o granicach (0,1104, 0,2268) z prawdopodobieństwem 95% zawiera nieznaną rzeczywistą wartość parametru 4.

Merytoryczna ocena: Wartość współczynnika jest zgodna z oczekiwaniami. Wyższy dochód rozporządzalny pozwala na zakup i utrzymanie samochodu.

X_7 – Liczba kradzieży samochodów

Interpretacja:

Wzrost liczby kradzieży o jednostkę zmniejsza liczbę zarejestrowanych samochodów o 0,001 na 1000 ludności.

Oceny parametru X_7 odchylają się przeciętnie o 0,0003 od jego nieznanej rzeczywistej wartości. Przedział o granicach (-0,0022, -0,0008) z prawdopodobieństwem 95% zawiera nieznaną rzeczywistą wartość.

Merytoryczna ocena: Choć wpływ jest niewielki, jest zgodny z intuicją, że wyższa liczba kradzieży może zniechęcać do posiadania samochodu. Niewielki wpływ może być spowodowany tym iż dane nie są podane wskaźnikiem.

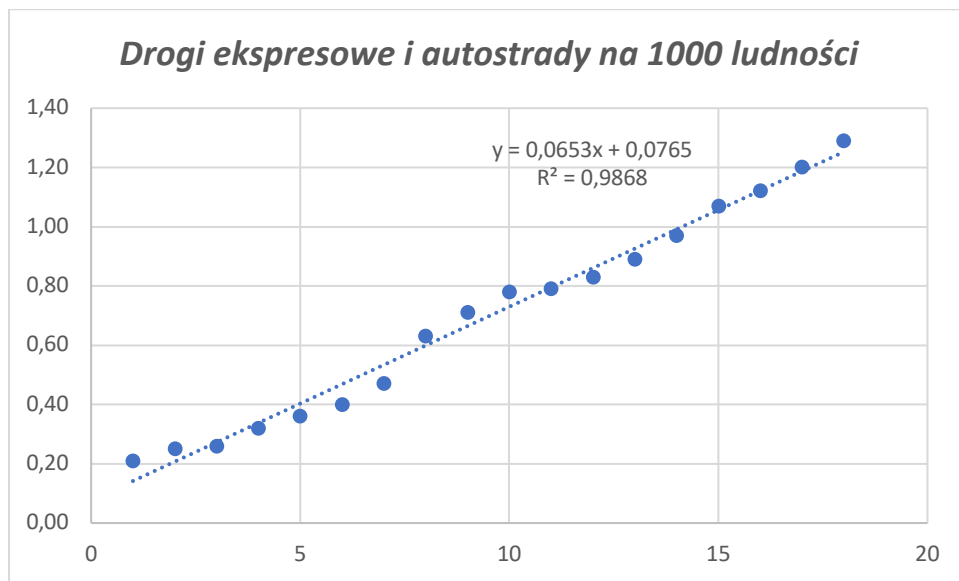
Podsumowanie:

Podsumowując, model jest solidny pod względem wysokiej wartości R^2 , niskiego wskaźnika zmienności i istotności zmiennych. Jednakże, wskazania na nieliniowość i potencjalne zagrożenie autokorelacji dodatkowo sugerują, aby poprawić ten model. Wprowadzić dopasowanie modeli nieliniowych, jak również zastanowić się nad zestawem zmiennych objaśniających może wtedy uniknąć się zagrożenia autokorelacją dodatnią.

Modele jednowymiarowe:

X_1 - Liniowy

$$\hat{Y} = 0,08 + 0,07X$$

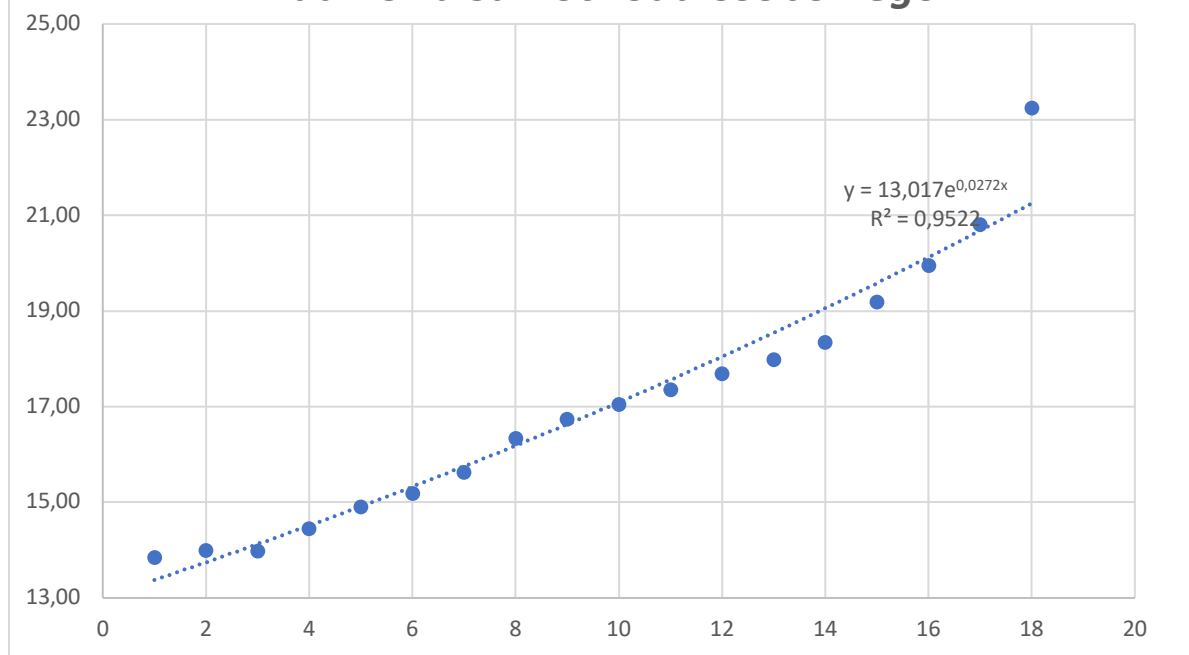


Prognoza na roku 2023 wynosi: 1,3179085

X_3 - Wykładniczy

$$\hat{Y} = 13,02e^{0,03X}$$

Cena przeciętnego mycia mechanicznego nadwozia samochodu osobowego

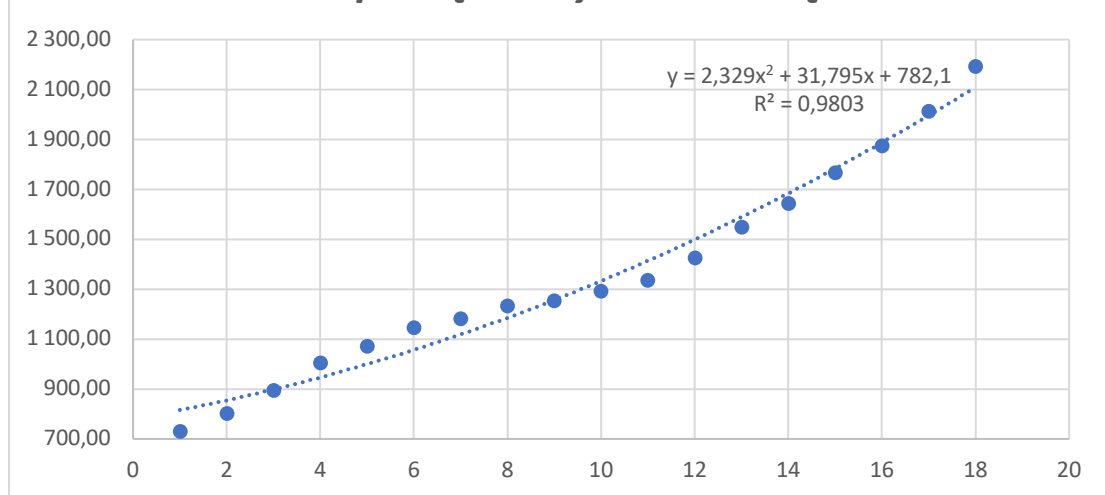


Prognoza na roku 2023 wynosi: 21,8340448

X₄ - Parabola

$$\hat{Y} = 782,1 + 31,8X + 2,33X^2$$

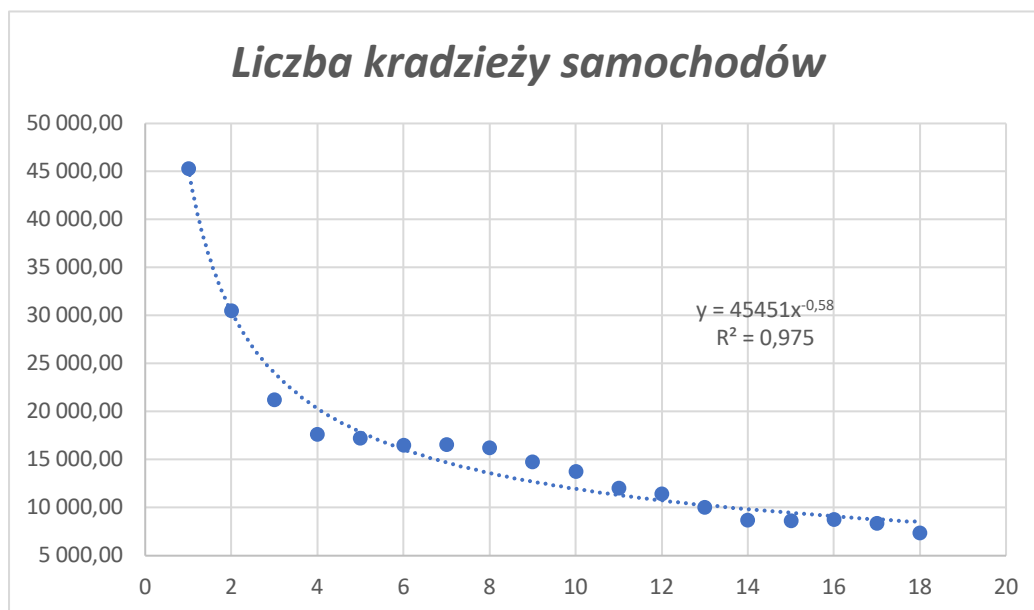
Przeciętny miesięczny dochód rozporządzalny na 1 osobę



Prognoza na roku 2023 wynosi: 2226,97363

X_7 - Potęgowa

$$\hat{Y} = 45451,37 X^{-0,58}$$



Prognoza na roku 2023 wynosi: 8237,970918

Prognoza liczby zarejestrowanych samochodów osobowych na 1000 ludności na rok 2023 wynosi 729,983094 czyli około 730 samochodów osobowych zarejestrowanych na 1000 ludności w Polsce będzie było w 2023 roku