



Imię i nazwisko
Hubert Dalmata

Tytuł w języku polskim:

Szacowanie emisji gazów cieplarnianych netto w tonach na mieszkańca, na terenie Polski w latach 2004-2021 na podstawie modelu ekonometrycznego.

Tytuł w języku angielskim:

Estimation of net greenhouse gas emissions in tons per capita in Poland from 2004 to 2021 based on an econometric model

Praca Zaliczeniowa Ekonometria

Promotor:brak.....

Kierunek: Informatyka i Ekonometria studia stacjonarne

Poznań 29.05.2024

W moim projekcie zająłem się badaniem wpływu wielu czynników na emisję gazów cieplarnianych netto. Wybrałem ten temat, gdyż emisja gazów cieplarnianych nierozdzielnie łączy się z postępującym na świecie globalnym ociepleniem oraz zmianami klimatu. Problemy te w niedalekiej przyszłości będą wpływały w znacznym stopniu na jakość mojego życia jak i moich rówieśników, dlatego postanowiłem zająć się tym tematem.

Y- emisja gazów cieplarnianych netto w tonach per capita

X1- PKB realne per capita (zmiana procentowa w stosunku do poprzedniego okresu), ponieważ PKB może wpływać na emisję gazów cieplarnianych poprzez różnorodne mechanizmy, włączając w to rozwój energii odnawialnej, zmiany w konsumpcji oraz stymulowanie innowacji, regulacje środowiskowe.

X2- Produkcja w przemyśle (zmiana procentowa w stosunku do poprzedniego okresu), ponieważ przy produkcji dóbr i usług oraz energii prawie zawsze wydzielają się znaczne ilości gazów cieplarnianych.

X3- Udział energii odnawialnej w końcowym zużyciu energii brutto w procentach według sektorów, ponieważ korzystanie ze źródeł energii odnawialnej powoduje mniejsze korzystanie z nieodnawialnych źródeł energii, co za tym idzie mniejszą emisyjność.

X4- linie kolejowe ogółem na 100 km², gdyż koleje są znacznie mniej emisyjne od samolotów czy samochodów.

X5- samochody osobowe na 1000 ludności, ponieważ mniejsza liczba samochodów oznacza, że więcej osób podróżuje środkami transportu zbiorowego, które są znacznie mniej emisyjne.

Model hipotetyczny

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

	Y	X1	X2	X3	X4	X5	
2004	9,3	5	1,121212121	6,882	6,5	292,5	
2005	9,3	3,6	1,050675676	6,867	6,5	323,4	
2006	9,8	6,2	1,123794212	6,859	6,5	351,1	
2007	10	7,1	1,094420601	6,903	6,4	382,7	
2008	9,8	4,2	1,020915033	7,686	6,5	421,6	
2009	9,4	1,9	0,960307298	8,676	6,5	432,2	
2010	9,9	2,8	1,110666667	9,281	6,5	447,4	
2011	9,7	5	1,073229292	10,337	6,5	470,3	
2012	9,5	1,5	1,011185682	10,955	6,4	486,4	

2013	9,3	0,9	1,026548673	11,452	6,2	503,7
2014	9,2	3,9	1,03125	11,605	6,2	519,9
2015	9,4	4,5	1,044932079	11,881	6,2	539,1
2016	9,5	3	1,031	11,396	6,1	564,0
2017	9,9	5,2	1,068865179	11,059	6,1	585,5
2018	9,9	6	1,058076225	14,936	6,2	610,0
2019	9,8	4,5	1,042881647	15,377	6,2	634,7
2020	9,3	-0,19	0,978618421	16,102	6,2	659,4
2021	10,1	7,5	1,148739496	15,613	6,2	682,4

DANE:

EUROSTAT: <https://ec.europa.eu/eurostat/web/main/data/database>

GŁÓWNY URZĘD STATYSTYCZNY: <https://bdl.stat.gov.pl/BDL/start>

Regresja krokowa -wstecz

Model 1: Estymacja KMNK, wykorzystane obserwacje 2004-2021 (N = 18)

Zmienna zależna (Y): Y

	<i>Współczynnik</i>	<i>Błąd stand.</i>	<i>t-Studenta</i>	<i>wartość p</i>	
const	-2,32109	3,06383	-0,7576	0,4633	
X1	0,0546139	0,0280985	1,944	0,0758	*
X2	1,85561	1,16848	1,588	0,1383	
X3	-0,124053	0,0423799	-2,927	0,0127	**
X4	1,31943	0,407189	3,240	0,0071	***
X5	0,00554986	0,00133207	4,166	0,0013	***
Średn.aryt.zm.zależnej	9,616667	Odch.stand.zm.zależnej	0,289523		
Suma kwadratów reszt	0,273129	Błąd standardowy reszt	0,150867		
Wsp. determ. R-kwadrat	0,808330	Skorygowany R-kwadrat	0,728468		
F(5, 12)	10,12154	Wartość p dla testu F	0,000556		
Logarytm wiarygodności	12,15274	Kryt. inform. Akaike'a	-12,30548		
Kryt. bayes. Schwarza	-6,963252	Kryt. Hannana-Quinna	-11,56886		
Autokorel.reszt - rho1	0,158266	Stat. Durbina-Watsona	1,603883		

ZMIENNA X2 wyrzucamy. Statystyka t-Studenta największa!!

Model 2: Estymacja KMNK, wykorzystane obserwacje 2004-2021 (N = 18)

Zmienna zależna (Y): Y

	Współczynnik	Błąd stand.	t-Studenta	wartość p	
const	-0,460170	2,99199	-0,1538	0,8801	
X1	0,0891399	0,0188133	4,738	0,0004	***
X3	-0,110170	0,0438286	-2,514	0,0259	**
X4	1,32558	0,430345	3,080	0,0088	***
X5	0,00508469	0,00137342	3,702	0,0027	***
Średn.aryt.zm.zależnej	9,616667	Odch.stand.zm.zależnej	0,289523		
Suma kwadratów reszt	0,330530	Błąd standardowy reszt	0,159453		
Wsp. determ. R-kwadrat	0,768049	Skorygowany R-kwadrat	0,696680		
F(4, 13)	10,76159	Wartość p dla testu F	0,000449		
Logarytm wiarygodności	10,43598	Kryt. inform. Akaike'a	-10,87196		
Kryt. bayes. Schwarza	-6,420096	Kryt. Hannana-Quinna	-10,25810		
Autokorel.reszt - rho1	-0,113860	Stat. Durbina-Watsona	2,203010		

POSTAĆ MODELU TEORETYCZNEGO

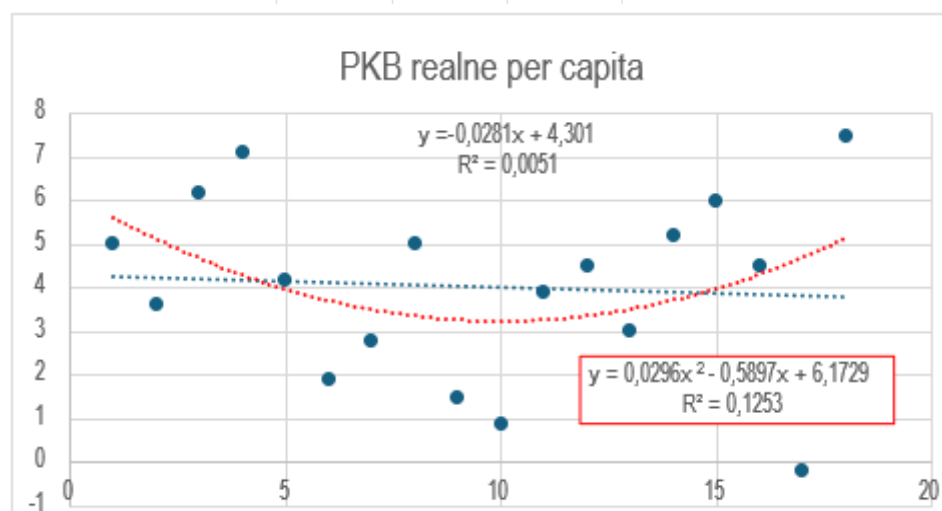
$$\hat{Y}=0,0050858*X_5+1,325577*X_4-0,11017*X_3+0,08914*X_1-0,046017$$

modele jednowymiarowe

X1 – PARABOLA (wybrałem parabolę gdyż ma największą wartość R^2 jak i najmniejszą wartość V z policzonych modeli)

X1	T	T^2
5	1	1
3,6	2	4
6,2	3	9
7,1	4	16
4,2	5	25
1,9	6	36
2,8	7	49
5	8	64
1,5	9	81
0,9	10	100
3,9	11	121
4,5	12	144
3	13	169
5,2	14	196
6	15	225
4,5	16	256
-0,19	17	289

0,029556	-0,58969	6,172941
0,020583	0,402463	1,6608
0,125339	2,092575	#N/D
1,074752	15	#N/D
9,412395	65,68303	#N/D
t1	3,716847	
t2	-1,4652	
t3	1,435967	
p1	0,002067	
p2	0,163511	NIEISTOTNE
p3	0,171538	NIEISTOTNE
V=	51,87%	ZA DUZE



7,5	18	324
-----	----	-----

POMOCNICZY
$\hat{Y} = b_1 + b_2X + b_3Z$

$$\hat{Y} = 6,1729 - 0,58969 * X + 0,029559 * Z$$

ORIGINALNY
$\hat{Y} = b_1 + b_2X + b_3X^2$

$$\hat{Y} = 6,1729 - 0,58969 * X + 0,029559 * X^2$$

$$W = \left(x_w = \frac{-b_2}{2b_3}; y_w = \frac{-\Delta}{4b_3} \right)$$

delta	-0,38206
Xw	9,975733
Yw	3,231644

INTERPRETACJA:

Maksymalną zmianę procentową PKB per capita wynoszącą 3,23% otrzymamy w przybliżeniu w 2011 roku.

PROGNOZA:

Można przypuszczać, że w 2022(t=19) zmiana procentowa PKB realnego per capita w stosunku do roku 2021 wyniesie : 5,638627451

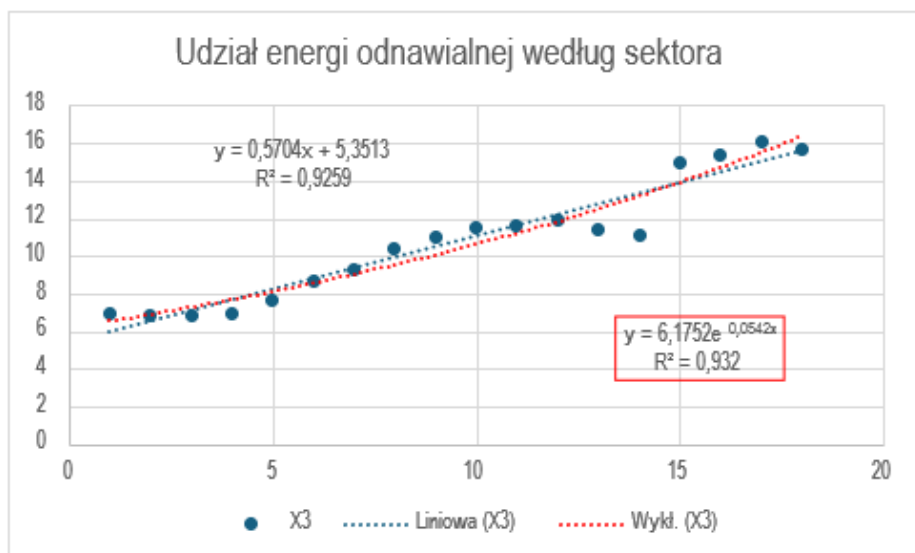
X3 –MODEL WYKŁADNICZY (wybrałem wykładniczy gdyż ma najmniejszą wartość V jak i największą wartość R² ze wszystkich modeli)

W=LnX	T
1,928909307	1
1,92672733	2
1,925561659	3
1,9319561	4
2,039400492	5
2,160560593	6
2,2279693	7
2,335729692	8
2,393795973	9
2,438164387	10
2,45143604	11
2,474940485	12

REGLINP:

0,054184	1,820548
0,003446	0,037306
0,939203	0,075862
247,1693	16
1,422455	0,09208
t1	48,80058
t2	15,72162
p1	7,75E-19
p2	3,77E-11
V=	3,25%

2,433262417	13
2,403244576	14
2,703774406	15
2,732872887	16
2,778943488	17
2,748103901	18



$$\hat{W} = b_1^* + b_2 X$$

$$\hat{W} = 1,820548 + 0,054184 * X$$

$$\hat{Y} = b_1 e^{b_2 X}$$

$$\hat{Y} = 6,175241 * e^{0,054184 * X}$$

b1*	1,820548
b2	0,054184
b₁ = e^{b₁*}	
b1	6,175241

INTERPRETACJA:

co roku procentowy udział energii odnawialnej według sektora rósł o 5,4184%
w roku 2003 (t=0) teoretyczna wartość procentowego udziału i odnawialnej według sektora wynosiła 6,175241

PROGNOZA:

Można przypuszczać, że w 2022(t=19) procentowy udział energii odnawialnej według sektora wyniesie: 17,28859278

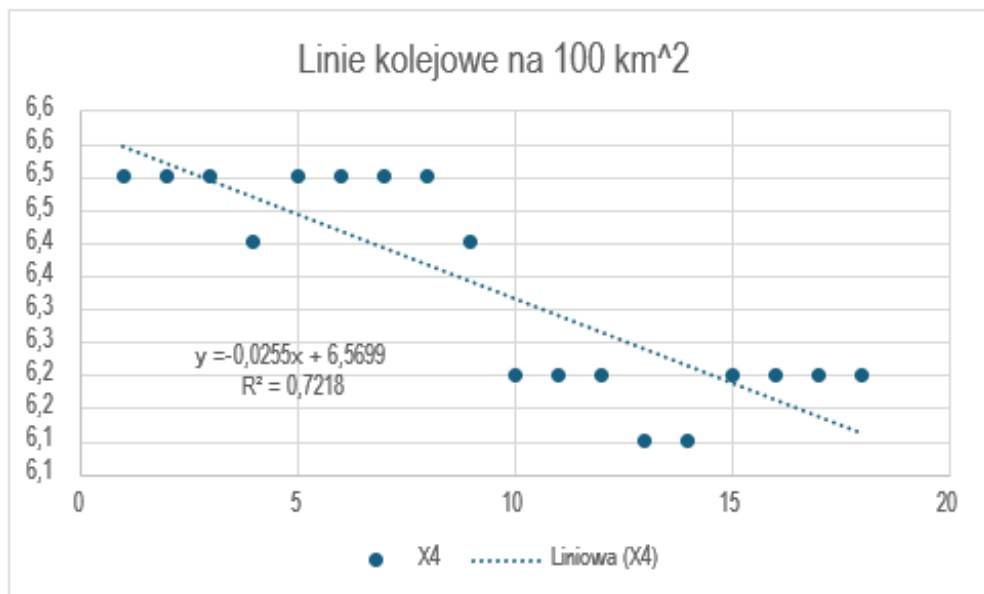
X4 –MODEL LINIOWY (wybrałem model liniowy gdyż jest najprostszy ,ma największą wartość R² oraz drugie co do wartości najmniejsze V z różnicą rzędu 0,5 %)

REGLINP:

X4	t
6,5	1
6,5	2
6,5	3
6,4	4
6,5	5
6,5	6
6,5	7
6,5	8
6,4	9

-0,02549	6,569935
0,003956	0,042819
0,721843	0,087073
41,52155	16
0,314804	0,121307
t1	153,4344
t2	-6,44372
p1	8,89E-27
p2	8,12E-06
V=	1,38%

6,2	10
6,2	11
6,2	12
6,1	13
6,1	14
6,2	15
6,2	16
6,2	17
6,2	18



$$\hat{Y} = b_0 + b_1 * X_4$$

$$\hat{Y} = 6,569935 - 0,02549 * X_4$$

INTERPRETACJA:

Co roku liczba linii kolejowych na 100 km² maleje o 0,02549

W 2003 liczba linii kolejowych wyniosła 6,569935

PROGNOZA:

Można przypuszczać, że w 2022 (t=19) liczba linii kolejowych na 100 km² wyniesie: 6,085620915

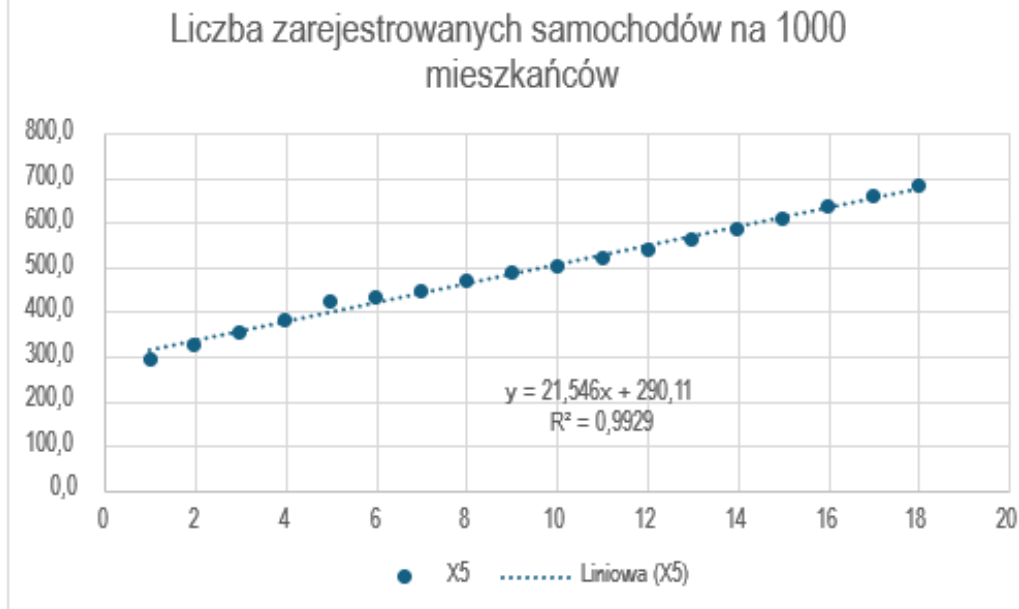
X5 –MODEL LINIOWY (wybrałem model liniowy gdyż jest najprostszy, ma zdecydowanie największą wartość R² oraz jedno z najmniejszych V z różnicą około 1%)

X5	t
292,5	1
323,4	2
351,1	3
382,7	4
421,6	5
432,2	6
447,4	7
470,3	8
486,4	9
503,7	10
519,9	11
539,1	12
564,0	13

REGLINP:

21,54561	290,1111
0,457042	4,947187
0,992852	10,06012
2222,313	16
224911,4	1619,296
t1	58,64163
t2	47,14141
p1	4,16E-20
p2	1,34E-18
V=	2,03%

585,5	14
610,0	15
634,7	16
659,4	17
682,4	18



$$\hat{Y} = b_0 + b_1 * X_5$$

$$\hat{Y} = 290,1111 + 21,54561 * X_5$$

INTERPRETACJA:

Co roku liczba zarejestrowanych samochodów na 1000 mieszkańców rośnie o 21,54561.

W 2003 roku liczba zarejestrowanych samochodów na 1000 mieszkańców wynosiła : 290,1111

PROGNOZA:

Można przypuszczać ,że w 2022(t=19) liczba zarejestrowanych samochodów na 1000 mieszkańców wyniesie: 699,4777778

testy sprawdzające założenia modelu

Test White'a na heteroskedastyczność reszt (zmiennosc wariacji resztowej)

Estymacja KMNK, wykorzystane obserwacje 2004-2021 (N = 18)

Zmienna zależna (Y): uhat^2

	współczynnik	błąd standardowy	t-Studenta	wartość p	
const	203,740	66,3870	3,069	0,0546	*
X1	1,04833	0,276214	3,795	0,0321	**
X3	1,60023	0,646626	2,475	0,0897	*
X4	-60,4646	19,5662	-3,090	0,0537	*
X5	-0,0902563	0,0274176	-3,292	0,0460	**
sq_X1	-0,000264003	0,00142579	-0,1852	0,8649	
X2_X3	-0,0123988	0,00551522	-2,248	0,1101	
X2_X4	-0,159216	0,0419918	-3,792	0,0322	**
X2_X5	0,000209898	0,000149706	1,402	0,2555	
sq_X3	-0,0130317	0,00574714	-2,268	0,1082	
X3_X4	-0,236582	0,0917470	-2,579	0,0819	*
X3_X5	0,000515880	0,000300874	1,715	0,1849	
sq_X4	4,49005	1,44752	3,102	0,0532	*
X4_X5	0,0133070	0,00396322	3,358	0,0438	**
sq_X5	-1,35406e-06	4,79768e-06	-0,2822	0,7961	

Uwaga: macierz danych jest osobiłwa!

Wsp. determ. R-kwadrat = 0,921841

Statystyka testu: $TR^2 = 16,593133$,
z wartością $p = P(\text{Chi-kwadrat}(14) > 16,593133) = 0,278509$

Hipoteza zerowa: heteroskedastyczność reszt nie występuje

WNIOSKI Możemy przyjąć założenie o homoskedastyczności w modelu (p -wartość = 0,2785 > 0,05).

Test Goldfelda-Quandt

0,005682	-0,07031	-0,1439	0,091588	8,621765	0,000526	-2,90787	0,066428	0,095166	25,99941
0,002015	1,451085	0,091955	0,041645	9,612404	0,004334	3,270314	0,14302	0,030525	20,65261
0,79642	0,168823	#N/D	#N/D	#N/D	0,863987	0,171005	#N/D	#N/D	#N/D
3,912079	4	#N/D	#N/D	#N/D	6,352255	4	#N/D	#N/D	#N/D
0,445995	0,114005	#N/D	#N/D	#N/D	0,743029	0,116971	#N/D	#N/D	#N/D
F empiryczne: 1,026019									
F kryt (0,1, T/2-K, T/2-K)= 4,10725					Fkryt>f empiryczne				
BRAK PODSTAW DO ODRZUCENIA HIPOTEZY O SKŁADNIKI LOSOWE SĄ HOMOSKEDASTYCZNE									

Test RESET na specyfikację -

Hipoteza zerowa: specyfikacja poprawna

Statystyka testu: $F(2, 11) = 1,66743$

z wartością $p = P(F(2, 11) > 1,66743) = 0,233076$

WNIOSKI. Nie mamy podstaw do odrzucenia hipotezy zerowej. Model ma postać liniową (p -wartość = 0,2331 > 0,05).

Test na normalność rozkładu reszt - Test Doornika-Hansena (1994)

Rozkład częstości dla residual, obserwacje 1-18

liczba przedziałów = 7, średnia = -5,92119e-016, odch.std. = 0,159453

Przedziały	średnia	liczba	częstość	skumulowana
< -0,22831	-0,27106	1	5,56%	5,56% *
-0,22831 - -0,14281	-0,18556	3	16,67%	22,22% *****
-0,14281 - -0,057313	-0,10006	2	11,11%	33,33% ***
-0,057313 - 0,028184	-0,014565	4	22,22%	55,56% *****
0,028184 - 0,11368	0,070933	4	22,22%	77,78% *****
0,11368 - 0,19918	0,15643	3	16,67%	94,44% *****
>= 0,19918	0,24193	1	5,56%	100,00% *

Hipoteza zerowa: składnik losowy ma cechy rozkładu normalnego

Statystyka testu: $\text{Chi-kwadrat}(2) = 0,186378$

z wartością $p = 0,911021$

WNIOSKI. Nie ma podstaw do odrzucenia hipotezy zerowej o normalnym rozkładzie składnika losowego ($p\text{-wartość} = 0,9110 > 0,05$).

TEST JBT	
T=	18
odch std	0,135509
S=	-0,18449
K=	2,239689
JBT=	0,059518
Chi kwadrat	5,991465
Istotność Chi	0,970679
Rozkład	JEST NORMALNY

TEST na autokorelację

Autokorel.reszt - rho1 -0,113860 Stat. Durбина-Watsona 2,203010

Statystyka testu Durбина-Watsona dla 5% poziomu istotności, $n = 18$, $k = 4$

$dL = 0,8204$

$dU = 1,8719$

składnik losowy nie wykazuje autokorelacji pierwszego rzędu

WNIOSKI. Nie ma podstaw do odrzucenia hipotezy zerowej o braku autokorelacji pierwszego rzędu w składniku losowym, ponieważ wartość DW jest większa od górnej granicy ($dU = 1,8719$).

(TEST NORMALNIE WYKONUJE SIE DLA $T \geq 30$ WIEC WYNIK JEST OBARCZONY DUŻYM BŁEDEM!!!!!!)

TEST LAGRANGEA

Test LM na autokorelację rzędu 4 -

Hipoteza zerowa: brak autokorelacji składnika losowego

Statystyka testu: LMF = 1,58902

z wartością $p = P(F(4, 9) > 1,58902) = 0,258578$

WNIOSKI. Nie mamy podstaw do odrzucenia hipotezy zerowej o braku autokorelacji 4 rzędu ($p\text{-wartość} = 0,2586 > 0,05$).

BADANIE DOBROCI

WSPÓŁCZYNNIK DETERMINACJI

$R^2 = 76,80\%$ Uzyskany model w 76,8% wyjaśnia jak kształtuje się emisja gazów cieplarnianych netto per capita w zależności od (PKB realne per capita, Procentowym Udziale energii odnawialnej w końcowym zużyciu energii, linii kolejowych na 100 km², liczbie zarejestrowanych samochodów na 1000 osób). **Adj. $R^2 = 70\%$**

WSPÓŁCZYNNIK INDETERMINACJI

$\phi^2 = 23,2\%$ emisji gazów cieplarnianych zależy od innych czynników, nieuwzględnionych w modelu.

OCENA WPŁYWU CZYNNIKÓW LOSOWYCH NA ZMIENNĄ OBJAŚNIANĄ

$S = 0,159453352$, Wartości empiryczne odchylają się od wartości teoretycznych przeciętnie o 0,1594 ton/ na mieszkańca

$V =$ Co stanowi 1,66% przeciętnych emisji gazów cieplarnianych netto per capita.

BŁĄD SZACUNKU PARAMETRU

$d1 = 2,991994057$, Oceny parametru pierwszego (wyraz wolny) mogą się odchyłać od jego nieznannej rzeczywistej wartości przeciętnie o 2,99.

$d2 = 0,018813324$, Oceny parametru drugiego mogą się odchyłać od jego nieznannej rzeczywistej wartości przeciętnie o 0,018813.

$d3 = 0,043828624$, Oceny parametru trzeciego mogą się odchyłać od jego nieznannej rzeczywistej wartości przeciętnie o 0,0438.

$d4 = 0,43034504$, Oceny parametru czwartego mogą się odchyłać od jego nieznannej rzeczywistej wartości przeciętnie o 0,43.

$d5 = 0,001373425$, Oceny parametru piątego mogą się odchyłać od jego nieznannej rzeczywistej wartości przeciętnie o 0,001.

WNIOSEKOWANIE O WEKTORZE PARAMETRÓW

$F = 10,76159477$

$F_{\alpha} = 3,179117053$

$F > F_{\alpha}$, wektor parametrów jest łącznie istotny

X5	X4	X3	X1	Wyraz wolny
0,005085	1,325577	-0,11017	0,08914	-0,46017
0,001373	0,430345	0,043829	0,018813	2,991994
0,768049	0,159453	#N/D	#N/D	#N/D
10,76159	13	#N/D	#N/D	#N/D
1,09447	0,33053	#N/D	#N/D	#N/D

PRZEDZIAŁY UFNOŚCI

$ud1 = -6,923980284$ $ug1 = 6,003640077$, Przedział o granicach -6,92233 i 6,00 z prawdopodobieństwem 0,95 zawiera nieznaną rzeczywistą wartość parametru 1 (wyraz wolny)

$ud2 = -6,923980284$ $ug2 = 6,003640077$, Przedział o granicach 0,048 i 0,12978 z prawdopodobieństwem 0,95 zawiera nieznaną rzeczywistą wartość parametru 2

$ud3 = -6,923980284$ $ug3 = 6,003640077$, Przedział o granicach -0,20486 i -0,01548 z prawdopodobieństwem 0,95 zawiera nieznaną rzeczywistą wartość parametru 3

$ud4 = -6,923980284$ $ug4 = 6,003640077$, Przedział o granicach 0,395873 i 2,255281 z prawdopodobieństwem 0,95 zawiera nieznaną rzeczywistą wartość parametru 4

$ud5 = -6,923980284$ $ug5 = 6,003640077$, Przedział o granicach 0,002118 i 0,0,008 z prawdopodobieństwem 0,95 zawiera nieznaną rzeczywistą wartość parametru 5

POSTAĆ MODELU TEORETYCZNEGO

$$\hat{Y} = 0,0050858 * X5 + 1,325577 * X4 - 0,11017 * X3 + 0,08914 * X1 - 0,046017$$

INTERPRETACJA

-Jeżeli Liczba samochodów zarejestrowanych na 1000 osób wzrośnie o 1 jednostkę to emisja gazów cieplarnianych wzrośnie o 0,005 ton /na mieszkańca, ceteris paribus. Sugeruje to, że większa liczba samochodów prowadzi do wyższych emisji gazów cieplarnianych, co jest zgodne z intuicją.

-Jeżeli Liczba linii kolejowych ogółem na 100 km² wzrośnie o 1 jednostkę to emisja gazów cieplarnianych netto wzrośnie o 1,325577 ton /na mieszkańca, ceteris paribus. To może sugerować, że rozwój infrastruktury kolejowej jest związany z wyższymi emisjami, być może z powodu energii używanej w transporcie kolejowym lub związanych z tym działań gospodarczych.

- Każdy 1% wzrost udziału energii odnawialnej w końcowym zużyciu energii brutto (przy założeniu, że inne zmienne pozostają stałe) jest związany ze spadkiem emisji gazów cieplarnianych netto o 0,11017 ton per capita. Sugeruje to, że większe wykorzystanie energii odnawialnej pomaga w redukcji emisji gazów cieplarnianych.

- Każdy 1% wzrost PKB realnego per capita (przy założeniu, że inne zmienne pozostają stałe) jest związany ze wzrostem emisji gazów cieplarnianych netto o 0,08914 ton per capita. Sugeruje to, że wzrost gospodarczy, mierzony jako wzrost PKB na mieszkańca, prowadzi do wzrostu emisji gazów cieplarnianych.

-WYRAZ WOLNY (NIE DO ZINTERPRETOWANIA)

WNIOSKI

Mój model wskazuje na to, że wzrost PKB realnego per capita, liczby samochodów osobowych oraz długości linii kolejowych zwiększa emisje gazów cieplarnianych, podczas gdy wzrost udziału energii odnawialnej zmniejsza te emisje (co sugeruje bardziej skomplikowane relacje między emisjami gazów cieplarnianych a długością linii kolejowych). Plusem jest to, że tylko jedna zmienna X_2 przy budowie modelu była nieistotna. Model przeszedł wszystkie testy z pozytywnymi rezultatami. Zależność między zmienną zależną a zmiennymi niezależnymi jest liniowa, reszty są homoskedastyczne, Składniki losowe nie są skorelowane, składnik losowy ma rozkład normalny. V =(współczynnik zmienności losowej) na poziomie 1,66% jest dobrym wynikiem. R^2 na poziomie 76,6% jest zdecydowanie do poprawy. Należało by dodać jeszcze jedną lub 2 zmienne istotne do modelu w celu poprawy R^2 .