



AKADEMIA GÓRNICZO-HUTNICZA
im. Stanisława Staszica w Krakowie



Wydział: Zarządzanie

Tytuł projektu: Budowa i analiza modelu ekonometrycznego dla
własnych danych

Przedmiot: Ekonometria

Autor: Luboński Jakub

Kierunek: Informatyka i Ekonometria

Rok: 2

09.06.2022 Kraków

Cel projektu:

Analiza wpływu wybranych czynników na liczbę zarejestrowanych samochodów osobowych w 2020 roku. Sprawdzenie czy przygotowany zestaw zmiennych (ilość wypadków drogowych, migracja ludności, wynagrodzenie, zgromadzone zapasy wspólnot mieszkaniowych oraz liczba osób bezrobotnych) może wyjaśniać liczbę zarejestrowanych samochodów osobowych. Odpowiedni dobór zmiennych, budowa modelu oraz analiza.

Hipotezy badawcze:

- Z zestawu przygotowanych zmiennych objaśniających jest możliwość znalezienia optymalnego podzbioru.
- Model dobrze opisuje wpływ na liczbę zarejestrowanych samochodów.

CZĘŚĆ I

Opis danych:

Wszystkie dane zostały pobrane z Głównego Urzędu Statystycznego. Dotyczą 2020 roku. Każda zmienna posiada 380 obserwacji, jedna obserwacja to jeden powiat. W skład nich wchodzi jedna zmienna objaśniana oraz sześć zmiennych objaśnianych.

Y – Samochody osobowe

Liczba zarejestrowanych samochodów osobowych na rok 2020.

X1 – Wypadki drogowe

Wypadek drogowy - zdarzenie mające związek z ruchem pojazdów na drogach publicznych, w wyniku którego nastąpiła śmierć lub uszkodzenie ciała osób. Za śmiertelną ofiarę wypadku drogowego uznaje się osobę zmarłą w wyniku doznanych obrażeń na miejscu lub w ciągu 30 dni. Za ranną ofiarę wypadku drogowego uznaje się osobę, która doznała obrażeń ciała i otrzymała pomoc lekarską.

X2 – Ilość dróg o nawierzchni twardej w przeliczeniu na 100 km².

X3 - Migracje na pobyt stały między powiatowe i zagraniczne wg ekonomicznych grup wieku

Współczynnik salda migracji: saldo migracji na pobyt stały ludności o określonym wieku na 10 tys. osób w tym wieku. Wskaźniki odnoszące się do liczby i struktury ludności (płeć, grupy wieku) od 2010 roku zostały przeliczone zgodnie z bilansem przygotowanym w oparciu o wyniki NSP 2011. Ze względu na niedostateczną jakość danych dotyczących migracji zagranicznych na pobyt stały za 2015 r., dane te, a także dane o migracjach długookresowych, nie zostały opublikowane.

X4 - Przeciętne miesięczne wynagrodzenia brutto.

Dane o wynagrodzeniach w ujęciu brutto, tj. łącznie z zaliczkami na poczet podatku dochodowego od osób fizycznych oraz od 1999 r. ze składkami na obowiązkowe ubezpieczenia społeczne (emerytalne, rentowe i chorobowe) płaconymi przez ubezpieczonego pracownika.

X5 - Bezrobotni zarejestrowani w ciągu roku .

Dane o liczbie bezrobotnych zarejestrowanych, obejmujących osoby zgodnie z ustawą z dnia 20 kwietnia 2004 r. o promocji zatrudnienia i instytucjach rynku pracy, obowiązującą od 1 kwietnia 2004 r. (Dz. U. z 2021 r. poz. 1100).

X6 – Koszty utrzymania zasobów.

Koszty utrzymania zasobów lokalowych dotyczą mieszkań zlokalizowanych w budynkach, które są w 100% własnością gminy, zakładu pracy, Skarbu Państwa, TBS oraz innych podmiotów mieszkaniowych. Koszty dotyczą wszystkich mieszkań, bez względu na status własności. Koszty utrzymania wykazywane są łącznie dla mieszkań i lokali użytkowych znajdujących się w budynkach mieszkalnych. Dane o kosztach utrzymania zasobów lokalowych przekazywane są przez jednostki, które posiadają 21 i więcej mieszkań, a w przypadku budynków objętych wspólnotami mieszkaniowymi 8 i więcej mieszkań w powiecie.

Statystyki opisowe:

Zmienna	Średnia	Mediana	Min	Maks	Odch.stand	Wsp. zmienności	Skośność	Kurtoza	Zakres Q3-Q1	Brakujące obs.
samochody	66089	49688	13838	1,43E+06	89010	1,3468	10,518	145,73	36770	0
wypadki	61,947	41	5	918	82,94	1,3389	5,8808	44,883	39,75	0
drogi	123,7	91,1	21,8	502,6	91,801	0,74213	1,6609	2,1631	73,675	0
migracja	-2,963	-102	-894	6551	545,78	184,19	6,9644	65,825	153	0
wynagrodzenie	4773,9	4638,2	3872,1	8920,4	581,3	0,12177	2,6006	11,396	538,16	0
bezrobocie	3528,3	2982	693	30177	2533,2	0,71797	4,8334	38,56	1951,3	0
zasoby	44024	12327	237,8	3,51E+06	1,98E+05	4,5034	14,71	247,44	22718	0

Tabela 1

Tabela 1 przedstawia statystyki opisowe zmiennych.

Samochody – zmienna ma drugi największy rozstęp danych ze wszystkich zmiennych co odbija się na duże odchylenie standardowe, największe w zestawieniu medianę i średnią oraz zakres między kwartylami. Współczynnik zmienności powyżej 1 świadczy o bardzo dużym zróżnicowaniu obserwacji. Ogromna kurtoza świadczy o bardzo dużej ilości próbek znajdujących się daleko od średniej. Dodatnia i bardzo duża skośność informuje, że prawe ramię rozkładu jest bardzo wydłużone. Brak obserwacji odstających.

Wypadki – zmienna ta ma najmniejsze dodatnie minimum z całego zestawu. Może się pochwalić również najmniejszym odchyleniem. Współczynnik zmienności powyżej 1 świadczy o bardzo dużym zróżnicowaniu obserwacji. Duża kurtoza świadczy o bardzo dużej ilości próbek znajdujących się daleko od średniej. Dodatnia i bardzo duża skośność informuje, że prawe ramię rozkładu jest bardzo wydłużone. Zarówno ona jak i Kurtoza nie są jednak aż tak duże jak w przypadku samochodów. Brak obserwacji odstających.

Drogi – zmienna ta przyjmuje wartości dodatnie między 21,8 a 501. Może się pochwalić najmniejszą różnicą między maksimum, a minimum. Współczynnik zmienności 0,74 świadczy o silnym zróżnicowaniu obserwacji. Najmniejsza kurtoza oraz skośność w zestawie informują, że dane są równomiernie oddalone od średniej, a prawe ramię rozkładu jest nieznacząco dłuższe. Brak obserwacji odstających.

Migracja – jako jedyna przyjmuje wartości ujemne. Wyróżnia się ogromnym współczynnikiem zmienności, który wynosi aż 184,19. Bardzo duża skośność informuje o wydłużeniu prawego ramienia rozkładu zmiennej. Duża kurtoza świadczy o bardzo dużej ilości próbek znajdujących się daleko od średniej. Średnia oraz mediana ujemne jako jedyne w zestawie.

Wynagrodzenie – współczynnik zmienności wynoszący 0,12 świadczy o bardzo małej zmienności próbek. Kurtoza na pozór duża jest jednak drugą najmniejszą w zestawie. Duża skośność informuje o wydłużeniu prawego ramienia rozkładu zmiennej. Mediana znajduje się zdecydowania bliżej wartości minimalnych. Brak prób odstających.

Bezrobocie – współczynnik zmienności wynoszący 0,72 świadczy o silnej zmienności próbek. Bardzo duże odchylenie standardowe świadczące o sporym oddaleniu obserwacji od średniej. Znaczna skośność oraz kurtoza. Średnia oraz mediana znajduje się zdecydowania bliżej wartości minimalnych. Brak prób odstających.

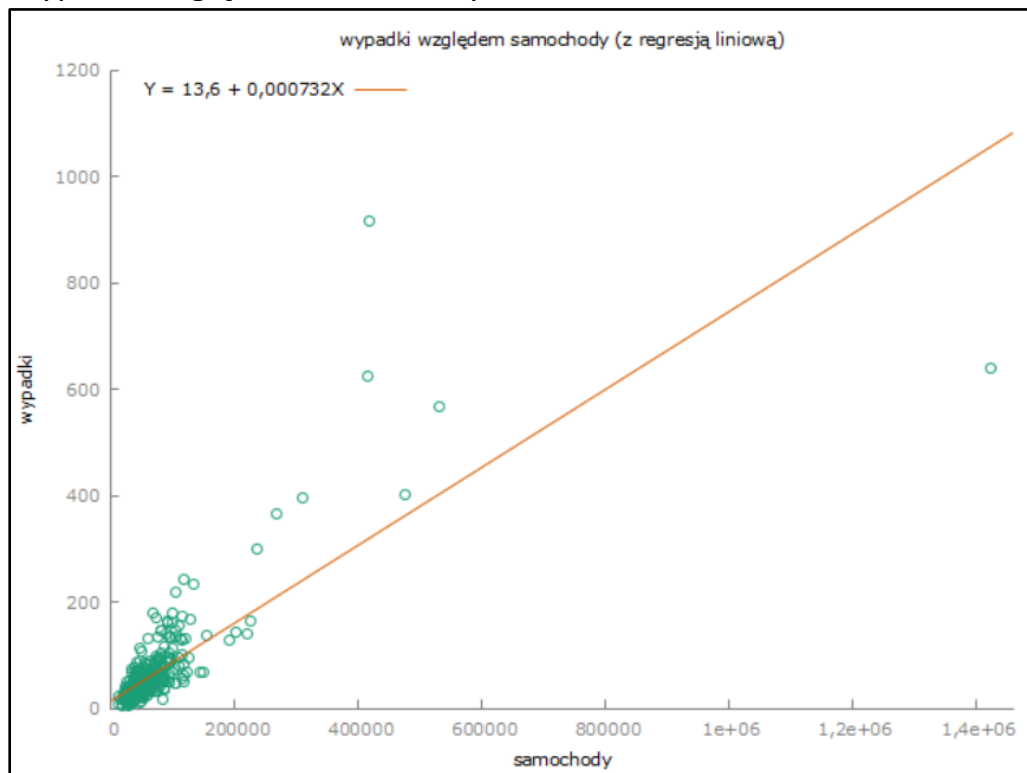
Zasoby – zmienna posiada największą wartość z całego zestawu, odchylenie standardowe, skośność oraz kurtozę. Świadczy to, o bardzo dużym oddaleniu obserwacji od średniej maksymalnie wydłużając prawe ramię rozkładu. Współczynnik zmienności wynoszący 4,5 informuje o bardzo silnej zmienności próbek. Brak wartości odstających.

Wniosek:

Z powodu bardzo małej wartości współczynnika zmienności dla zmiennej wynagrodzenie powinna zostać usunięta i nie brana pod uwagę w dalszym badaniu.

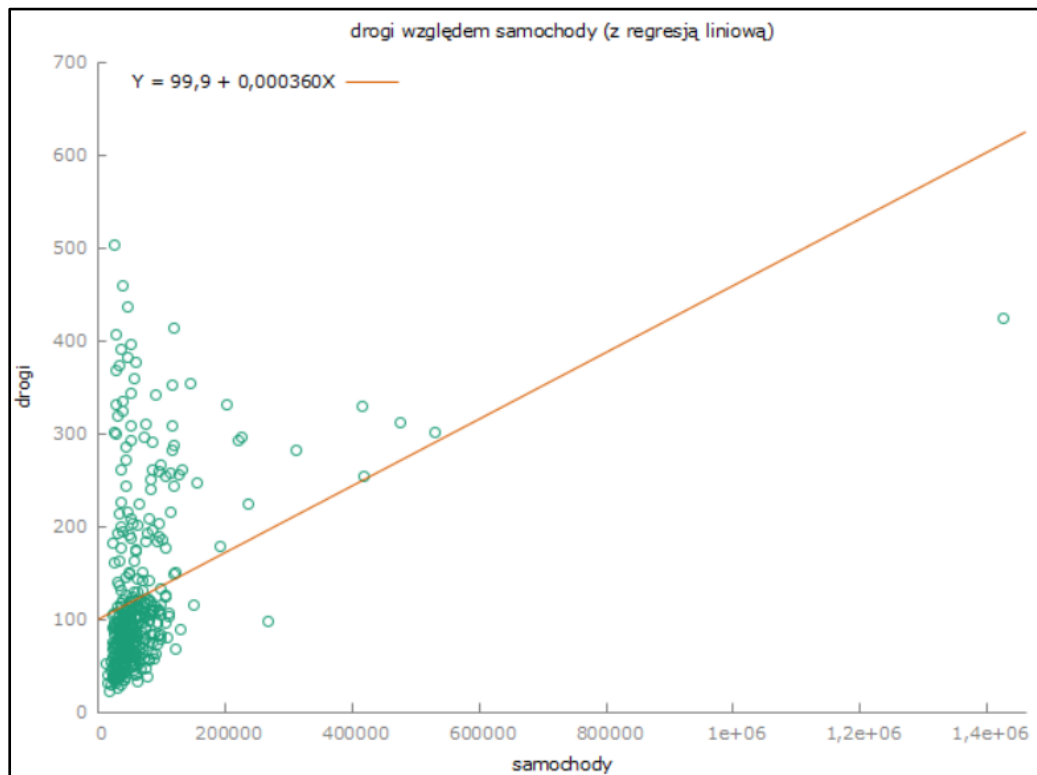
Wykresy zależności:

1. Wypadki względem samochodu.



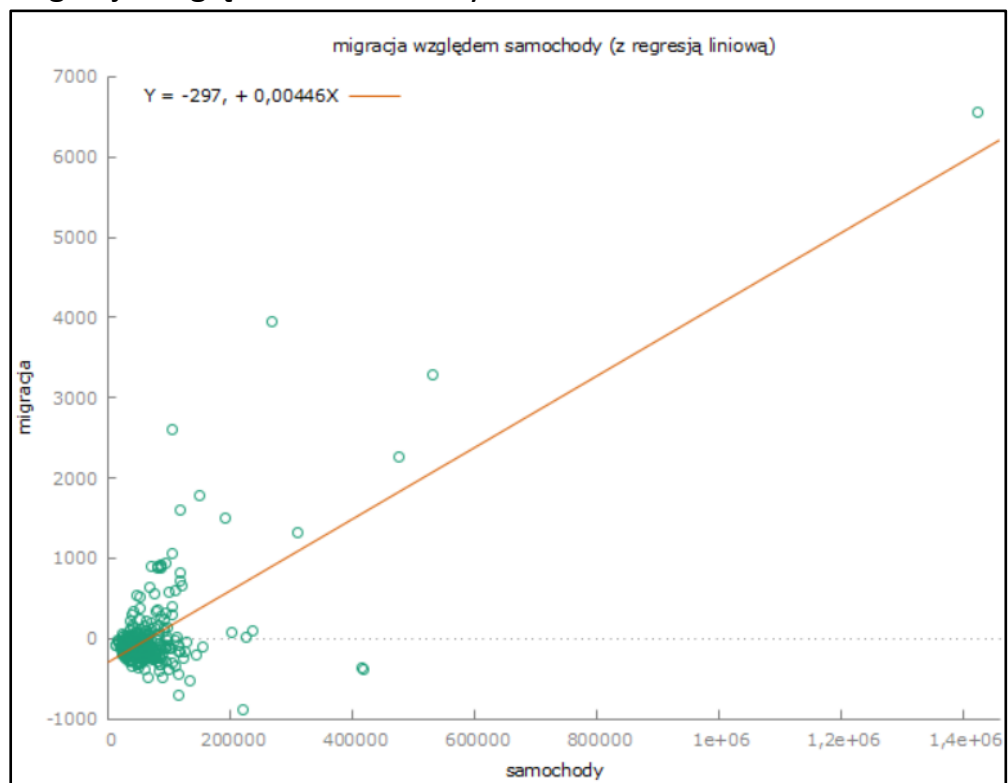
Wykres 1

2. Drogi względem samochodu



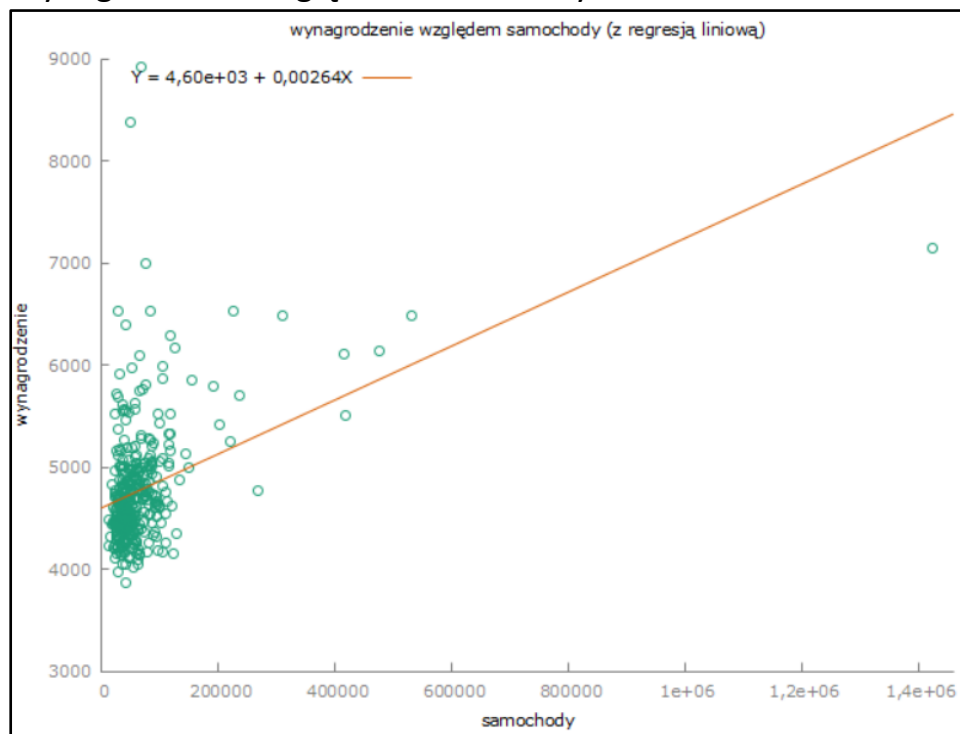
Wykres 2

3. Migracja względem samochodu.



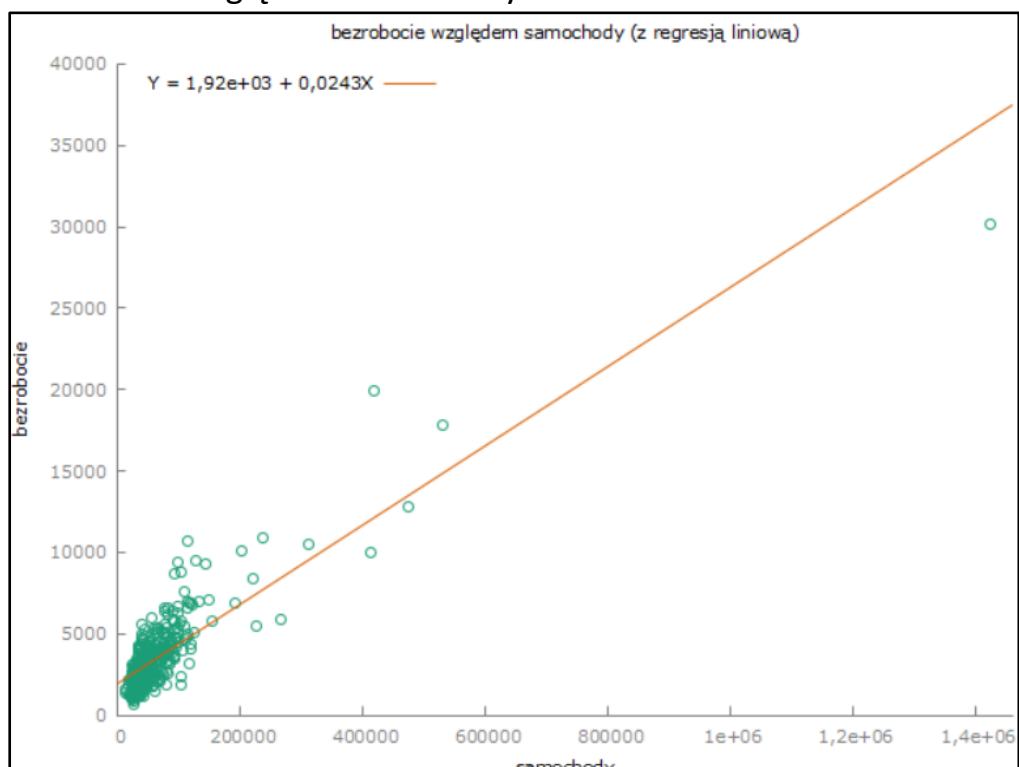
Wykres 3

4. Wynagrodzenie względem samochodu.



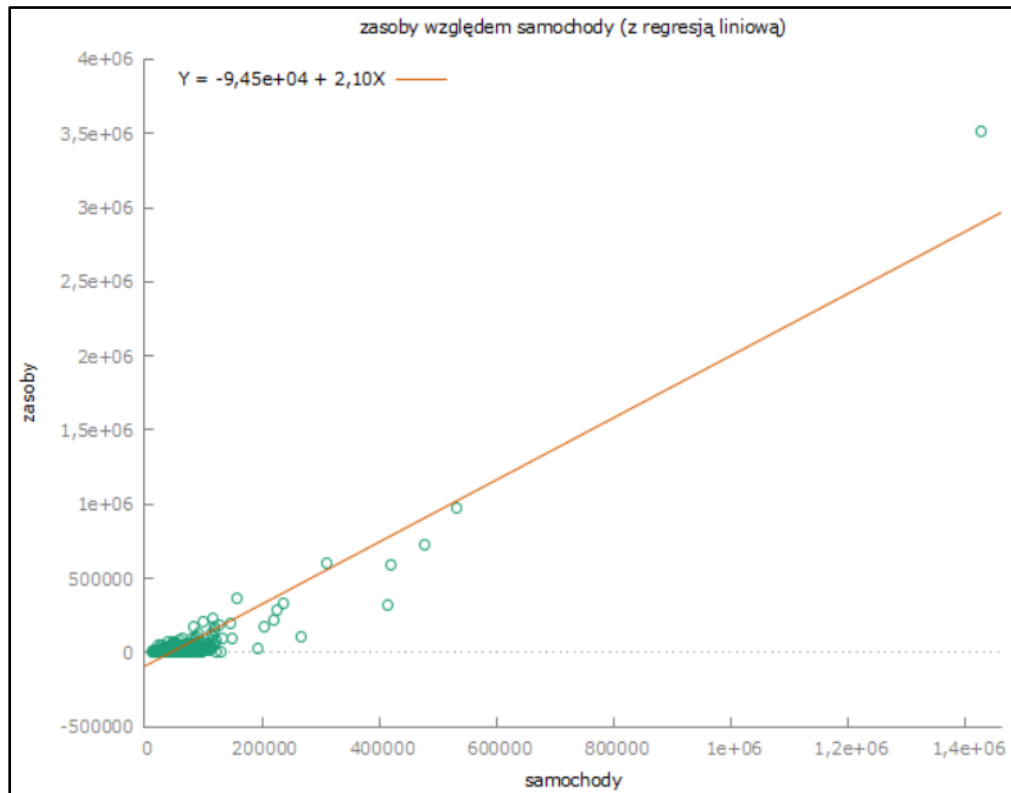
Wykres 4

5. Bezrobocie względem samochodu.



Wykres 5

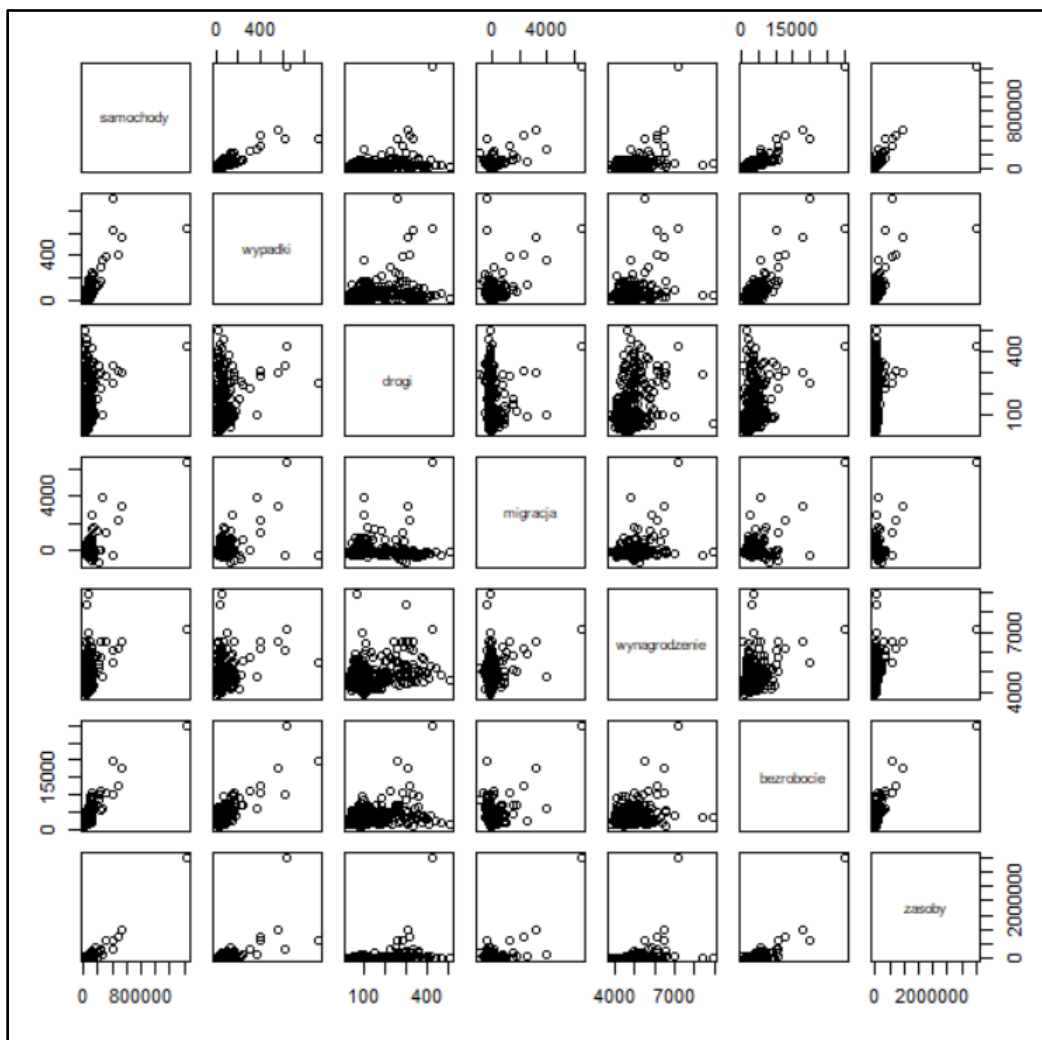
6. Zasoby względem samochodu.



Wykres 6

Na wszystkich wykresach można odnotować dodatnią zależność zmiennych oraz skupienie w dolnym lewym rogu wykresu.

7. Wykresy zależności między wszystkimi zmiennymi.



Wykres 7

Macierz korelacji między zmiennymi:

samochody	wypadki	drogi	migracja	wynagrodzenie	bezrobocie	zasoby	
1	0,7854	0,3487	0,7266	0,4049	0,8556	0,9407	samochody
	1	0,3048	0,4967	0,3414	0,7951	0,6297	wypadki
		1	0,1171	0,361	0,3334	0,3196	drogi
			1	0,3052	0,4827	0,6992	migracja
				1	0,2946	0,3749	wynagrodzenie
					1	0,7477	bezrobocie
						1	zasoby

Tabela 2

Z tabeli 2 można odczytać, że najmniejsze korelacje występują przy zmiennych wynagrodzenie oraz drogi. Potwierdza się więc potrzeba usunięcia zmiennej wynagrodzenie.

CZĘŚĆ II

Budowa modelu:

Stworzenie modelu za pomocą klasycznej metody najmniejszych kwadratów, w którym zmienna objaśniana to samochody, a zmiennymi objaśniającymi są wypadki, drogi, migracja, bezrobocie, wynagrodzenie oraz zasoby. Wykorzystanie do tego komendy `ols` w programie Gretl, o której informacje zostały zaczerpnięte z książki Using Gretl for Principles of Econometrics(Lee C. Adkins, 2012).

```
ols samochody const wypadki drogi migracja bezrobocie zasoby wynagrodzenie
```

Tabela 3 ze wstępnymi informacjami o modelu, którą wyświetlił program po utworzeniu modelu.

	współczynnik	błąd standardowy	t-Studenta	wartość p
const	-2827,55	8055,3	-0,3510	0,7258
wypadki	196,452	17,8708	10,99	1,54e-024 ***
drogi	18,6666	10,6326	1,756	0,0800 *
migracja	21,8161	2,30917	9,448	3,90e-019 ***
bezrobocie	7,6567	0,677873	11,3	1,21e-025 ***
zasoby	0,248882	0,0082885	30,03	1,44e-101 ***
wynagrodzenie	3,46271	1,6973	2,04	0,0420 **

Tabela 3

Wartości p-value w tabeli 3 przy zmiennych drogi oraz wynagrodzenie wynoszące odpowiednio 0,080 oraz 0,042 są stosunkowo wysokie i sugerują ich nieistotność przy objaśnianiu zmiennej zależnej samochody.

Poprawa modelu 1:

Wykorzystanie metod doboru(redukcji) zmiennych objaśniających:

- Metoda krokowa nie wykazała żadnych potencjalnych zmiennych do usunięcia
- Metoda Hellwiga(Porządkowanie liniowe obiektów metodą Hellwiga i Topsis – analiza porównawcza, Andrzej Bąk, 2016) pokazała, że najlepszy zbiór zmiennych objaśniających to: wypadki, migracja, bezrobocie oraz zasoby.

Na podstawie wyników wykorzystania metody Hellwiga(oraz wcześniejszych przestaneek wynikając z macierzy korelacji, wysokich wartości p-value oraz niskiego współczynnika zmienności z modelu zostały usunięte zmienne drogi oraz wynagrodzenie.

Nowy model:

```
ols samochody const wypadki migracja bezrobocie zasoby
```

	współczynnik	błąd standardowy	t-Studenta	wartość p
const	15399,7	1895,43	8,125	6,57e-015 ***
wypadki	205,48	17,8081	11,54	1,45e-026 ***
migracja	21,3015	2,29506	9,281	1,37e-018 ***
bezrobocie	7,59997	0,680971	11,16	3,62e-025 ***
zasoby	0,254605	0,00817297	31,15	4,50e-106 ***

Tabela 4

Wartości p-value w tabeli 4 dla każdej zmiennej w modelu są bardzo małe(poniżej 0,01) co sugeruje ich istotność w modelu.

Wykonanie kolejnych sprawdzeń na prawidłowość doboru zmiennych w modelu z wykorzystaniem automatycznych funkcji w programie Gretl:

- Test na normalność reszt.

Hipoteza zerowa: składnik losowy ma rozkład normalny.

Statystyka testu: Chi-kwadrat(2) = 185,163 z wartością p = 6,20082e[^](-41).

Wartość p jest dużo mniejsza niż 0,05, zatem odrzucamy hipotezę zerową.

Składnik losowy modelu nie ma rozkładu normalnego.

- Test White'a na heteroskedastyczność reszt (zmienność wariancji resztowej).

Hipoteza zerowa: heteroskedastyczność reszt nie występuje.

Statystyka testu: LM = 271,014 z wartością p = P(Chi-kwadrat(14) > 271,014) = 1,27079e[^](-49).

Wartość p jest dużo mniejsza niż 0,05, zatem odrzucamy hipotezę zerową.

W modelu występuje heteroskedastyczność.

- Test na nieliniowość (kwadraty).

Hipoteza zerowa: zależność jest liniowa.

Statystyka testu: LM = 43,114 z wartością $p = P(\text{Chi-kwadrat}(4) > 43,114) = 9,79922\text{e-}09$.

Wartość p jest dużo mniejsza niż 0,05, zatem odrzucamy hipotezę zerową.
Zależność nie jest liniowa.

- Test na nieliniowość (logarytmy).

Hipoteza zerowa: zależność jest liniowa

Statystyka testu: LM = 18,1296 z wartością $p = P(\text{Chi-kwadrat}(3) > 18,1296) = 0,000413591$.

Wartość p jest dużo mniejsza niż 0,05, jednak różnica nie jest już aż tak duża jak w poprzednich testach. Odrzucenie hipotezy zerowej.
Zależność nie jest liniowa.

- Test na dodane zmienne (wynagrodzenie, drogi).

Hipoteza zerowa: parametry regresji dla zmiennych drogi, wynagrodzenie są równe zero.

Statystyka testu: $F(2, 373) = 5,00396$, wartość p 0,00716884.

Pominięcie zmiennych poprawiło 1 z 3 kryteriów informacyjnych (AIC, BIC, HQC).
Model mógłby być gorszy po ponownym dodaniu zmiennych. Usunięcie z nich było dobrym posunięciem, jednak model dalej wymaga poprawy.

Poprawa modelu 2:

Zamiana wszystkich zmiennych na ich logarytmy. Redukcja obserwacji do 79.

```
ols l samochody const l wypadki l migracja l bezrobocie l zasoby
```

	współczynnik	błąd standardowy	t-Studenta	wartość p
const	4,6778	0,421267	11,1	2,02e-017 ***
l_wypadki	0,273985	0,0569277	4,813	7,66e-06 ***
l_migracja	0,0409166	0,0190044	2,153	0,0346 **
l_bezrobocie	0,497665	0,0788379	6,313	1,83e-08 ***
l_zasoby	0,114852	0,0251508	4,567	1,94e-05 ***

Tabela 5

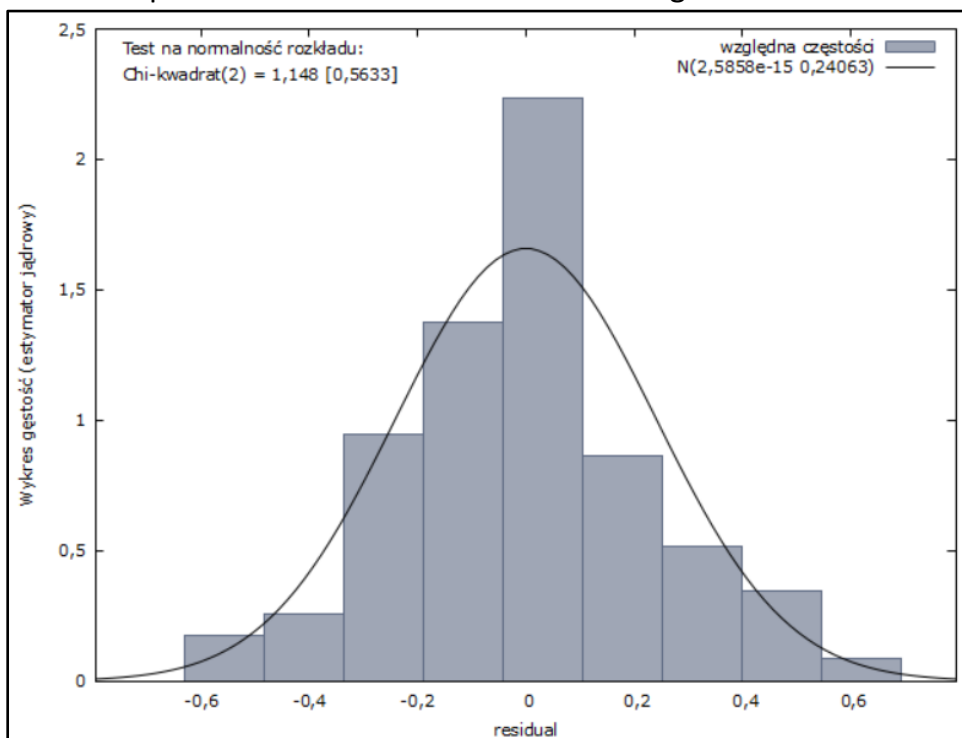
Wykonanie testów sprawdzających poprawność modelu.

- Test na normalność reszt.

Hipoteza zerowa: składnik losowy ma rozkład normalny

Statystyka testu: Chi-kwadrat(2) = 1,14776 z wartością $p = 0,563335$

Wartość p jest większa niż 0,05. Nieodrzućenie hipotezy zerowej. Składnik losowy modelu ma rozkład normalny. Wykres 8 przedstawia kształtowanie się reszt oraz podobieństwo do rozkładu normalnego.



Wykres 8

- Test White'a (Podstawy statystyki i ekonometrii Część II, Sabina Nowak, 2019) na heteroskedastyczność reszt (zmiennność wariancji resztowej).

Hipoteza zerowa: heteroskedastyczność reszt nie występuje

Statystyka testu: LM = 13,7024 z wartością $p = P(\text{Chi-kwadrat}(14) > 13,7024) = 0,472109$.

Wartość p jest dużo większa niż 0,05. Nieodrzućenie hipotezy zerowej.

W modelu nie występuje heteroskedastyczność.

- Test na nieliniowość (kwadraty).

Hipoteza zerowa: zależność jest liniowa.

Statystyka testu LM = 15,692 z wartością $p = P(\text{Chi-kwadrat}(4) > 15,692) = 0,00346155$.

Wartość p jest niewiele, ale mniejsza niż 0,05.

- Test na nieliniowość (logarytmy).

Hipoteza zerowa: zależność jest liniowa.

Statystyka testu: LM = 4,4064 z wartością $p = P(\text{Chi-kwadrat}(3) > 4,4064) = 0,220793$.

Wartość p jest większa niż 0,05. Nieodrzućmy hipotezy zerowej. Zależność jest liniowa.

- Testy na pominięcie zmiennych w modelu (dla każdej z osobna).

Hipoteza zerowa: parametr regresji dla zmiennej $I_wypadki$ jest równy zero.

Statystyka testu: $F(1, 74) = 23,1637$

z wartością $p = P(F(1, 74) > 23,1637) = 7,66391e-06$.

Hipoteza zerowa: parametr regresji dla zmiennej $I_bezrobocie$ jest równy zero.

Statystyka testu: $F(1, 74) = 39,8478$

z wartością $p = P(F(1, 74) > 39,8478) = 1,82532e-08$.

Hipoteza zerowa: parametr regresji dla zmiennej I_zasoby jest równy zero.

Statystyka testu: $F(1, 74) = 4,63543$

z wartością $p = P(F(1, 74) > 4,63543) = 0,0345775$.

Hipoteza zerowa: parametr regresji dla zmiennej I_zasoby jest równy zero.

Statystyka testu: $F(1, 74) = 20,8533$

z wartością $p = P(F(1, 74) > 20,8533) = 1,93667e-05$.

We wszystkich testach na pominięcie zmiennej wartości p_value są mniejsze niż 0,05, zatem hipoteza zerowa została odrzucona. Pominięcie jakiegokolwiek zmiennej objaśniającej nie ulepszyłoby modelu.

Prawie wszystkie testy wyszły pozytywnie dla modelu. W teście na nieliniowość kwadratową wartość p-value była tylko nie wiele mniejsza niż 0,05. Model z logarytmami zmiennych objaśniających l_migracja, l_wypadki, l_bezrobocie, l_zasoby oraz zmienną objaśnianą l_samochody jest zatem najlepszym rozwiązaniem z pozostałych i zostaje wybrany do pełnej analizy.

CZĘŚĆ IV

Ostateczny model

	współczynnik	błąd standardowy	t-Studenta	wartość p
const	4,6778	0,421267	11,1	2,02e-017 ***
l_wypadki	0,273985	0,0569277	4,813	7,66e-06 ***
l_migracja	0,0409166	0,0190044	2,153	0,0346 **
l_bezrobocie	0,497665	0,0788379	6,313	1,83e-08 ***
l_zasoby	0,114852	0,0251508	4,567	1,94e-05 ***

Tabela 6

Średn.aryt.zm.zależnej	11,27643	
Suma kwadratów reszt	4,284824	
Wsp. determ. R-kwadrat	0,886136	
F(4, 74)	143,975	
Odch.stand.zm.zależnej	0,694587	
Błąd standardowy reszt	0,24063	
Skorygowany R-kwadrat	0,879982	
Wartość p dla testu F	4,12E-34	

Tabela 7

Postać modelu:

Z tabeli 6 można odczytać, że utworzony model ma postać:

$$\log(y) = 4,68 + 0,27 \times \log(X_1) + 0,04 \times \log(X_2) + 0,5 \times \log(X_3) + 0,12 \times \log(X_4)$$

Interpretacja modelu:

- wzrost wartości zmiennej wypadki o jedną jednostkę spowoduje wzrost zmiennej zależnej o 0,27%

- wzrost wartości zmiennej migracja o jedną jednostkę odpowiada wzrostowi liczby zarejestrowanych samochodów o 0,04%
- wzrost wartości zmiennej bezrobocie o jedną jednostkę odpowiada wzrostowi liczby zarejestrowanych samochodów o 0,5%
- wzrost wartości zmiennej zasoby o jedną jednostkę odpowiada wzrostowi liczby zarejestrowanych samochodów o 0,11%.

Analiza:

Informacje o modelu uzyskane dzięki użyciu testów pomocnych przy doborze zmiennych:

1. Normalność składnika losowego.
2. Brak heteroskedastyczności reszt.
3. Model liniowo zależny dzięki testom nieliniowości.
4. Optymalny dobór zmiennych dzięki, między innymi, metodzie Hellwiga.
5. Istotność parametrów poprzez testy dodanych oraz pominiętych zmiennych.

Dalsze badania:

6. Współczynnik determinacji R^2 .

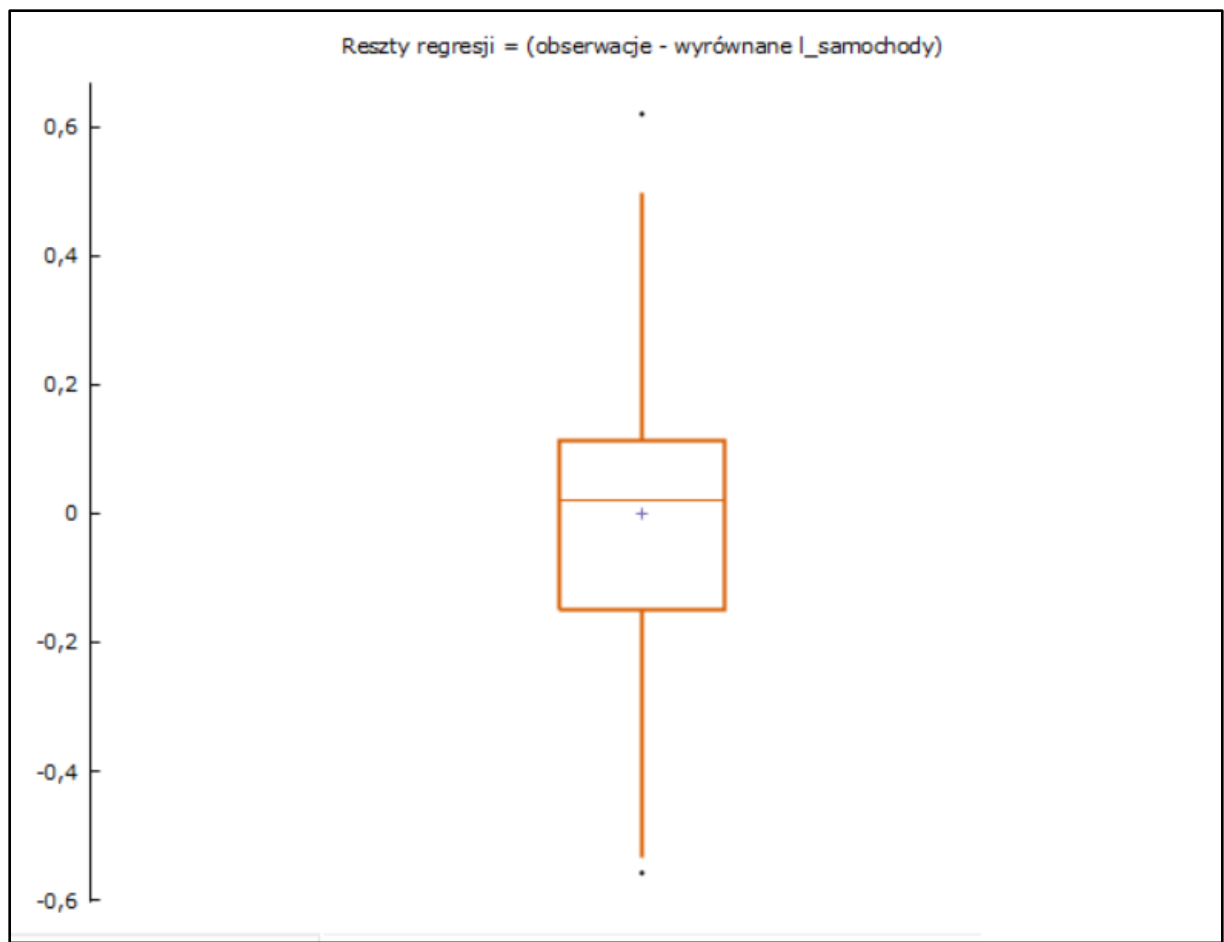
Wartość współczynnika determinacji R^2 widoczna w tabeli 7 informuje, że model umożliwia objaśnienie zmienności zmiennej Y w 88,6%. Taki wynik uważa się za dobre dopasowanie.

Zaletą jest zbliżona wartość skorygowanego współczynnika determinacji wynosząca 88%.

7. Efekt katalizy.

Natężenie efektu katalizy jest różnicą pomiędzy współczynnikiem determinacji R^2 a integralną pojemnością informacyjną zestawu zmiennych objaśniających. Obliczone natężenie wyniosło zaledwie 0,1, a algorytm nie wykrył żadnych katalizatorów.

8. Wartości odstające dla zmiennej objaśnianej $I_{\text{samochody}}$ (wykres pudełkowy).



Wykres 9

Z wykresu 9 można odnaleźć dwie wartości odstające.

9. Test t-Studenta na istotność pojedynczych parametrów.

Wartość krytyczna odczytana z tablicy rozkładu t-Studenta (liczba stopni swobody $n - (k + 1) = 75$) wynosi: 1,9944.

Wartości statystyki rozkładu t-Studenta widoczne w tabeli 6 są wyższe dla każdej zmiennej w modelu, zatem wszystkie zmienne objaśniające są istotne.

10. Test liczby serii.

Liczba serii (R) dla zmiennej 'e' = 38

Test niezależności oparty na liczbie dodatnich i ujemnych serii.

Hipoteza zerowa: próba jest losowa, dla R odpowiednio $N(40,5, 4,41588)$,

test z-score = -0,566139, przy dwustronnym obszarze krytycznym
 $p = 0,5713$.

Wartość p-value jest duża stąd brak zastrzeżeń w zakresie losowości próby oraz poprawności postaci modelu.

11. Test Ramsey'a na stabilność postaci analitycznej modelu.

Statystyka testu: $F = 3,165669$,
z wartością $p = P(F(2,72) > 3,16567) = 0,0481$.

Wartość p-value w przybliżeniu do jednej setnej wynosi 0,05, zatem można uznać, że postać wybranego modelu jest stabilna.

12. Test Chowa.

Kryterium podziału: połowa obserwacji.

Hipoteza zerowa: parametry modelu regresji w obu podpróbach są jednakowe.

Test Chowa na zmiany strukturalne przy podziale próby w obserwacji 39

$F(5, 69) = 0,207061$ z wartością $p = 0,9585$.

Wartość p-value jest większe niż 0,05. Brak podstaw do odrzucenia hipotezy zerowej. Parametry w obu podgrupach są jednakowe. Wynik wskazuje na poprawność dobrania liniowego modelu.

13. Współliniowość.

Zmiana zmiennych objaśnianych oraz obliczenie dla modeli współczynników determinacji R^2 .

```
ols l_wypadki const l_migracja l_bezrobocie l_zasoby l_samochody  
R1=$rsq  
ols l_migracja const l_wypadki l_bezrobocie l_zasoby l_samochody  
R2=$rsq  
ols l_bezrobocie const l_wypadki l_migracja l_zasoby l_samochody  
R3=$rsq  
ols l_zasoby const l_wypadki l_migracja l_bezrobocie l_samochody  
R4=$rsq
```

$R1 = 0,7104$

$R2 = 0,317778$

$R3 = 0,7973$

$R4 = 0,571676$

Przyjmuje się, że mamy do czynienia ze współliniowością, gdy R^2 przekracza 90%. Żaden z wyliczonych nie jest blisko tej wartości, zatem brak podstaw do istnienia współliniowości.

Ocena współliniowości za pomocą polecenia VIF.

Ocena współliniowości VIF(j) - czynnik rozdęcia wariancji.

VIF (Variance Inflation Factors) - minimalna możliwa wartość = 1.0.

Wartości > 10.0 mogą wskazywać na problem współliniowości, czyli tzw. rozdęcia wariancji.

l_migracja	1,465
l_bezrobocie	4,828
l_zasoby	2,259
l_samochody	6,689

Tabela 8

Dla żadnej zmiennej wartości obliczonego współczynnika w tabeli 8 nie przekracza 10. Brak podstaw do istnienia współliniowości.

14. Koincydencja.

Model jest koincydentny, jeżeli dla każdej zmiennej objaśniającej spełniony jest warunek: $sgn\ r_i = sgn\ \alpha_i$ (Ekonometria, Mieczysław Sobczyk, 2013).

l_samochody	l_wypadki	l_migracja	l_bezrobocie	l_zasoby	
1	0,8009	0,5626	0,7869	0,6594	l_samochody
	1	0,4793	0,6188	0,5192	l_wypadki
		1	0,4891	0,4127	l_migracja
			1	0,5527	l_bezrobocie
				1	l_zasoby

Tabela 9

W tabeli 10 przedstawiającej macierz korelacji między zmiennymi widać, że wszystkie współczynniki r_i są dodatnie. W modelu wszystkie parametry przy zmiennych objaśniających są również dodatnie, zatem model jest koincydentny.

15. Prognoza.

a) Prognoza punktowa dla wartości średnich.

```
ols l_samochody const l_wypadki l_migracja l_bezrobocie l_zasoby --vcv --quiet  
matrix x_tau = {1, mean(l_wypadki), mean(l_migracja), mean(l_bezrobocie), mean(l_zasoby)}  
matrix a = $coeff  
scalar prognoza_punktowa_dla_l_sam = x_tau*a
```

Prognoza punktowa dla wartości średniej zmiennej l_samochody wynosi 11,0106.

Wariancja prognozy.

```
scalar wariancja_prognozy_dla_l_sam= x_tau*$vcv*x_tau'+$ess/$df
```

Wariancja wynosi 0,0592554.

Błąd prognozy

```
scalar blad_prognozy_dla_l_sam=sqrt(wariancja_prognozy_dla_l_sam)
```

Błąd wynosi 0,243424. O tyle średnio oszacowana zmienna prognozowanej odchyła się od rzeczywistej wartości.

Prognoza punktowa dla wartości średniej zmiennej samochody.

```
scalar prognoza_punktowa_dla_sam = exp(prognoza_punktowa_dla_l_sam+0.5*wariancja_prognozy_dla_l_sam)
```

Prognoza punktowa dla wartości średniej zmiennej samochody wynosi 62333,1.

b) Prognoza przedziałowa

Prognoza z 95% przedziałem ufności dla zmiennej l_samochody.

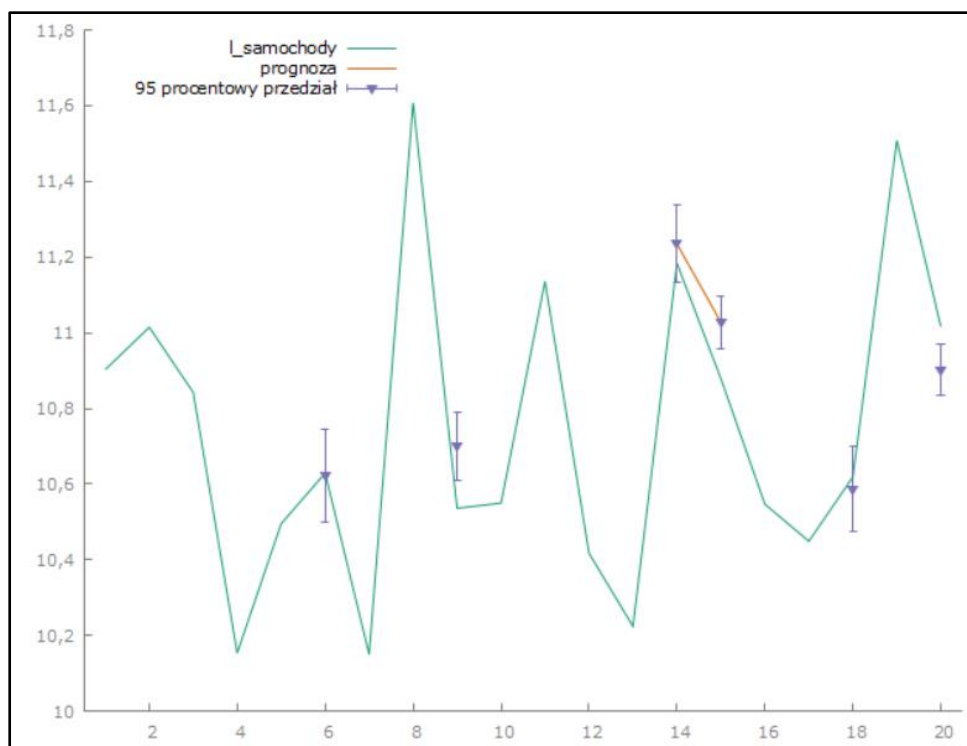
```
scalar dolna = prognoza_punktowa_dla_l_sam-critical(t,$df,0.025)*blad_prognozy_dla_l_sam  
scalar gorna = prognoza_punktowa_dla_l_sam+critical(t,$df,0.025)*blad_prognozy_dla_l_sam
```

Przedział, w którym 95% prawdopodobieństwem znajdą się przyszłe wartości zmiennej l_samochody to (10,5256 ; 11,4957).

Prognoza z 95% przedziałem ufności dla zmiennej samochody.

```
scalar dolna_dla_sam = exp(dolna)  
scalar gorna_dla_sam = exp(gorna)
```

Przedział, w którym 95% prawdopodobieństwem znajdą się przyszłe wartości zmiennej l_samochody to (37256,7 ; 98287,7).



Wykres 10

Wykres 10 przedstawia prognozę wartości średniej zmiennej objaśnianej $l_samochody$ dla 20 następnych obserwacji.

CZĘŚĆ V

PODSUMOWANIE:

Budowa właściwego modelu ekonometrycznego nie była łatwa, lecz ostatecznie udało się znaleźć postać, która pozytywnie przeszła wszystkie testy statystyczne na poprawność i stabilność modelu oraz istotność parametrów.

W trakcie badań okazało się, że średnie wynagrodzenie oraz ilość dróg o twardej nawierzchni nie mają statystycznego istotnego wpływu na liczbę zarejestrowanych samochodów w 2020 roku. Pozostałe zmienne (zasoby, migracje, bezrobocie, wypadki) dobrze objaśniają zmienną Y. By tak się stało, należało je przetransferować do postaci logarytmicznej. Model w takiej postaci pomyślnie przeszedł całą analizę aż do wyznaczenia prognozy, zarówno punktowej jak i przedziałowej. Wszystkie hipotezy badawcze zostały potwierdzone.