

Analiza cen dotyczących sprzedaży używanych samochodów w Wielkiej Brytanii

Spis treści

Spis treści	2
Wstęp.....	5
1. Statystyka opisowa zmiennych	6
1.1. Statystyka opisowa zmiennej Cena [zł]	6
1.2. Statystyka opisowa zmiennej Przebieg [km]	8
1.3. Statystyka opisowa zmiennej Rok rejestracji	9
1.4. Statystyka opisowa zmiennej Liczba poprzednich właścicieli	11
1.5. Statystyka opisowa zmiennej Pojemność silnika	13
1.6. Statystyka opisowa zmiennej Liczba drzwi.....	15
1.7. Statystyka opisowa zmiennej Liczba miejsc siedzących.....	17
1.8. Statystyka opisowa zmiennej Norma emisji spalin [Euro].....	19
2. Macierz korelacji Pearsona	21
2.1. Macierz korelacji Pearsona ze zmiennymi z badania.....	21
3. Regresja.....	22
3.1. Wynik funkcji Regresja	22
3.2. Statystyki regresji	22
3.3. Podsumowanie wyników regresji wielorakiej	23
4. Testy	24
5. Wnioski	31

Wstęp

Prezentowana analiza dotyczy zachowania się ceny samochodu w zależności od różnych zmiennych. Analiza została wykonywana w programie Microsoft Excel 2021 z dodatkiem Analiza Danych. W badaniu wykorzystano zbiór danych z informacjami o sprzedaży używanych samochodów na terenie Wielkiej Brytanii. Zbiór danych został pobrany ze strony internetowej kaggle.com. Zawierał on 3654 rekordów. Po zaimportowaniu danych z pliku CSV do programu Excel dokonano niezbędnego czyszczenia danych. Czyszczenie danych obejmowało uzupełnianie brakujących wartości, korektę ewentualnych błędów i zmianę ceny auta z £ (funty) na zł (złotówki). Podobny proces wykonano przy zmianie przebiegu auta z mil na kilometry. Brakujące dane w kolumnach Pojemność silnika, Liczba drzwi, Liczba miejsc siedzących, Norma emisji spalin [Euro] uzupełniono ręcznie bazując na danych zdobytych w Internecie. Danych, których brakowało w kolumnie Przebieg [km] uzupełniono wartością średniej, którą obliczono z pozostałych wartości.

1. Statystyka opisowa zmiennych

Funkcja Statystyka opisowa służy do wyświetlenia podstawowych miar statystycznych. Do wyznaczenia statystyk opisowych posłużono się narzędziem Statystyka opisowa w zakładce Analiza danych w programie MS Excel. W tym celu należało wybrać zakładkę Dane → Analiza → Analiza danych → Statystyka opisowa. W okienku należało wybrać zakres wejściowy, czyli kolumnę z danymi do obliczeń. Następnie należało ewentualnie zaznaczyć opcję Tytuły w pierwszym wierszu. Na koniec należało zaznaczyć opcję Statystyki podsumowujące oraz wybrać miejsce, w którym statystyki mają się wygenerować.

1.1. Statystyka opisowa zmiennej Cena [zł]

Tabela 1 Miary statystyczne zmiennej Cena [zł]

Miary statystyczne	Cena[zł]
Średnia	29854,88588
Błąd standardowy	383,7532453
Mediana	20640
Tryb	12874,2
Odchylenie standardowe	23197,24091
Wariancja próbki	538111985,8
Kurtoza	1,8819304
Skośność	1,376579386
Zakres	172860
Minimum	2064
Maksimum	174924
Suma	109089753
Licznik	3654
Współczynnik zmienności	0,776999818

Źródło: Wynik funkcji Statystyka opisowa w programie MS Excel

Cena [zł] – cena samochodu w złotych za jaką był wystawiony na sprzedaż. Jest to zmienna ilościowa. Jest to też zmienna objaśniana, czyli zmienna, którą objaśniamy w modelu.

Średnia cena samochodu wynosi 29854,89 zł. Błąd standardowy wynosi 383,75 zł co oznacza, że gdyby przeprowadzane były próby losowe z tej samej populacji i obliczane by były wartości średniej dla każdej próby, to średnie z tych prób mogłyby się różnić od rzeczywistej średniej o 383,75 zł. Mediana, czyli wartość środkowa, wynosi 20 640 zł. Tryb zwraca najczęściej występującą wartość. Najczęściej występującą ceną w zbiorze danych jest 12874,20 zł. Odchylenie standardowe wynosi 23197,24 zł, oznacza to, że większość cen samochodów w badanym zbiorze danych różni się od średniej ceny o 23197,24 zł. Wartość wariancji jest to odchylenie standardowe podniesione do kwadratu. Kurtoza określa jak wiele obserwacji jest zbliżona do wartości średniej oraz czy większość z uzyskanych wyników ma wartość podobną

do średniej. W badanym przypadku kurtoza wynosi 1,88 i jest większa od 0, co oznacza, że jest więcej wartości odstających niż w rozkładzie normalnym oraz rozkład ceny samochodu jest wysmukły w porównaniu do rozkładu normalnego. Skośność wynosi 1,38 i świadczy to o prawostronnej asymetrii rozkładu. Zakres, czyli różnica między najwyższą, a najniższą ceną, wynosi 172 860 zł. Najniższa cena samochodu wynosi 2 064zł. Najwyższa cena samochodu wynosi 174 924 zł. Łączna suma cen wszystkich samochodów wynosi 109089753 zł. W badaniu uwzględniono 3654 rekordy. Współczynnik zmienności wynosi 78%. Informuje nas o silnej zmienności i oznacza, że jest duże zróżnicowanie cechy.

1.2. Statystyka opisowa zmiennej Przebieg [km]

Tabela 2 Miary statystyczne zmiennej Przebieg [km]

Miary statystyczne	Przebieg[km]
Średnia	131255,1936
Błąd standardowy	1052,560573
Mediana	128800
Tryb	128800
Odchylenie standardowe	63625,52365
Wariancja próbki	4048207260
Kurtoza	125,684335
Skośność	5,197616529
Zakres	1787259,39
Minimum	1,61
Maksimum	1787261
Suma	479606477,4
Licznik	3654
Współczynnik zmienności	0,484746713

Źródło: Wynik funkcji Statystyka opisowa w programie MS Excel

Przebieg [km] – przebieg samochodu w kilometrach z jakim był wystawiony na sprzedaż. Jest to zmienna ilościowa. Zmienna mówi ile kilometrów zostało przejechane samochodem przez wszystkich poprzednich właścicieli.

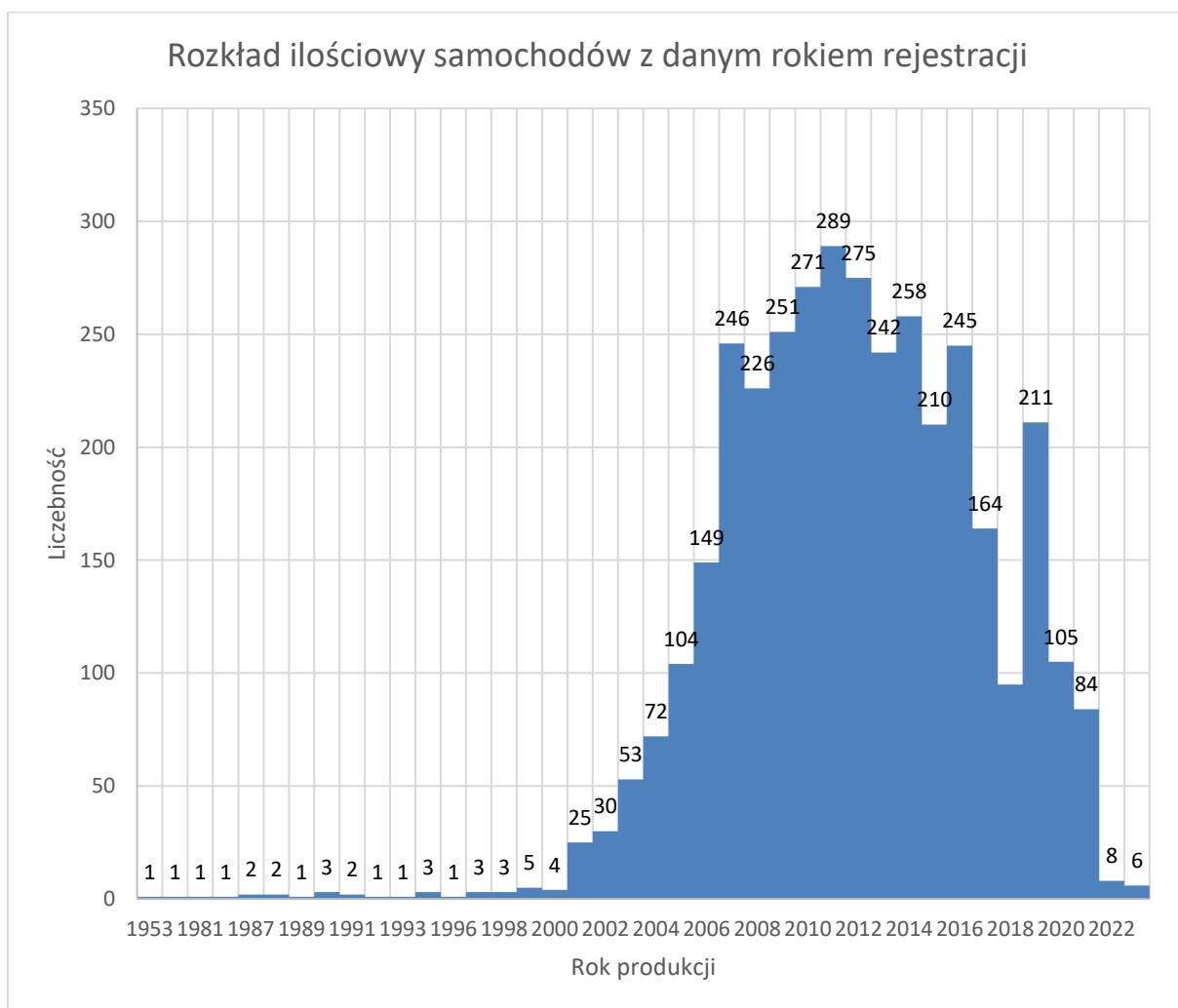
Średnia przebiegów samochodów wynosi 131255,19 km. Błąd standardowy wynosi 1052,56 km co oznacza, że gdyby przeprowadzane były próby losowe z tej samej populacji i obliczane by były wartości średniej dla każdej próby, to średnie z tych prób mogłyby się różnić od rzeczywistej średniej o 1052,56 km. Mediana, czyli wartość środkowa, wynosi 128800 km. Tryb zwraca najczęściej występującą wartość. Najczęściej występującym przebiegiem w zbiorze danych jest 128800 km. Odchylenie standardowe wynosi 63625,52 km, oznacza to, że większość przebiegów samochodów w badanym zbiorze danych różni się od średniego przebiegu o 63625,52 km. Wartość wariancji jest to odchylenie standardowe podniesione do kwadratu. Kurtoza określa jak wiele obserwacji jest zbliżona do wartości średniej oraz czy większość z uzyskanych wyników ma wartość podobną do średniej. W badanym przypadku kurtoza wynosi 125,68 i jest większa od 0, co oznacza że jest więcej wartości odstających niż w rozkładzie normalnym oraz rozkład przebiegu jest wysmukły w porównaniu do rozkładu normalnego. Skośność wynosi 5,20 i świadczy to o prawostronnej asymetrii rozkładu. Zakres wynosi 1787259,40 km. Najniższy przebieg wynosi 1,61 km. Najwyższy przebieg wynosi 1787261 km. Łączna suma przebiegów wszystkich samochodów wynosi 479606477,4 km. W badaniu uwzględniono 3654 rekordy. Współczynnik zmienności wynosi 48%. Informuje nas o silnej zmienności i oznacza, że jest duże zróżnicowanie cechy.

1.3. Statystyka opisowa zmiennej Rok rejestracji

Tabela 3 Miary statystyczne zmiennej Rok rejestracji

Miary statystyczne	Rok rejestracji
Średnia	2011,813082
Błąd standardowy	0,084475542
Mediana	2012
Tryb	2011
Odchylenie standardowe	5,106405019
Wariancja próbki	26,07537222
Kurtoza	6,533011345
Skośność	-0,925550536
Zakres	70
Minimum	1953
Maksimum	2023
Suma	7351165
Licznik	3654
Współczynnik zmienności	0,00253821

Źródło: Wynik funkcji Statystyka opisowa w programie MS Excel



Rysunek 1 Rozkład ilościowy z danym rokiem rejestracji

Źródło: opracowanie własne

Na wykresie można zauważyć, że w danym zbiorze danych najwięcej samochodów zostało pierwszy raz zarejestrowanych od lat 2007 do 2019. W latach 1953 do 2000 liczba rejestracji była bardzo mała i wynosiła od 1 do 3 zarejestrowanych samochodów w tych latach. Podobnie ma się zakres lat od 2022 do 2023.

Rok rejestracji – Rok pierwszej rejestracji samochodu. Mniejsza wartość oznacza, że samochód jest starszy. Jest to zmienna ilościowa.

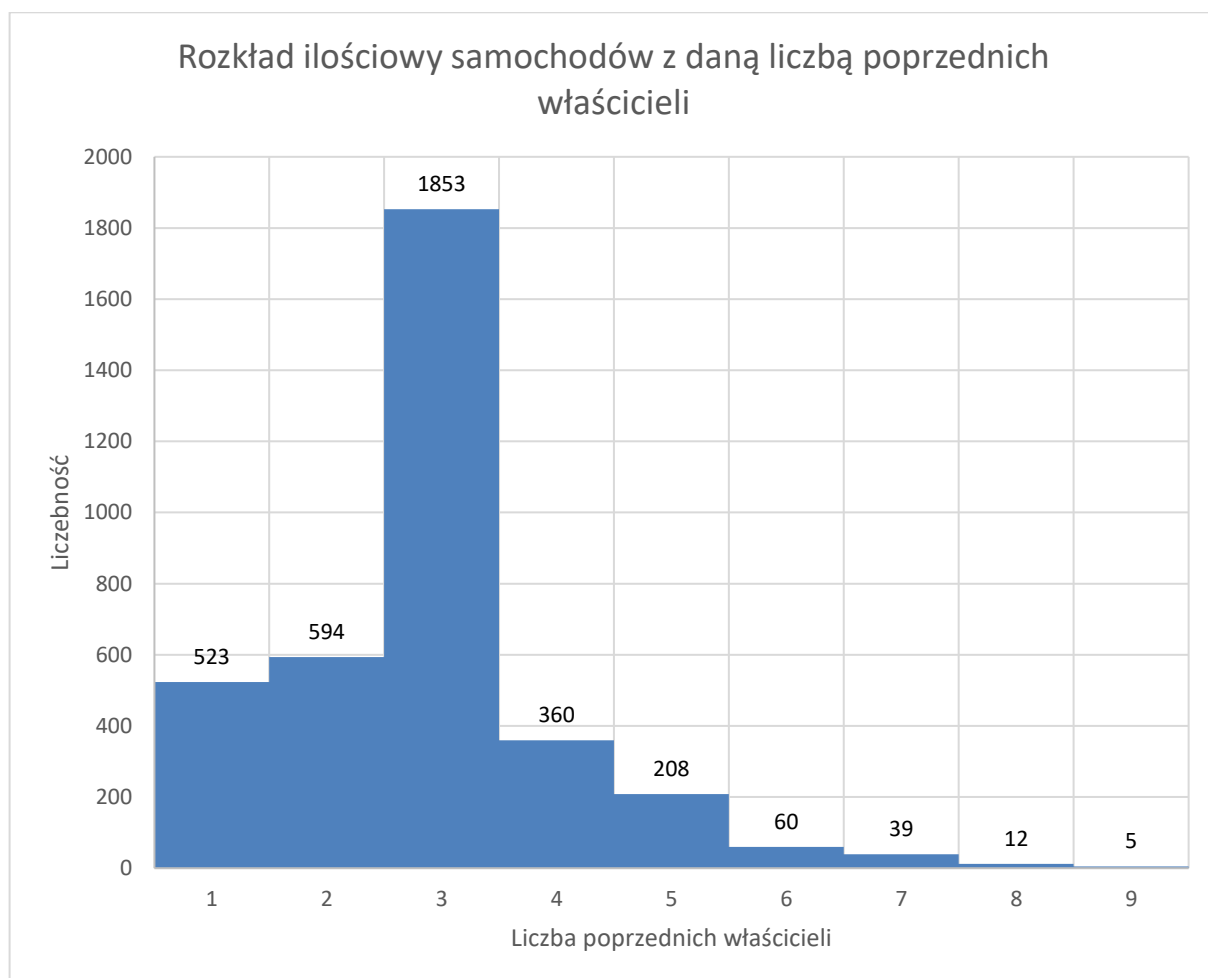
Średni rok rejestracji samochodu wynosi 2011. Błąd standardowy wynosi 0,08 co oznacza, że gdyby przeprowadzane były próby losowe z tej samej populacji i obliczane by były wartości średniej dla każdej próby, to średnie z tych prób mogłyby się różnić od rzeczywistej średniej o 0,08. Mediana, czyli wartość środkowa wynosi 2012. Tryb zwraca najczęściej występującą wartość. Najczęściej występującą wartością roku rejestracji w zbiorze jest 2011. Odchylenie standardowe wynosi 5,11 lat, oznacza to że większość roków rejestracji w badanym zbiorze różni się od średniego roku rejestracji o 5,11 lat. Wartość wariancji jest to odchylenie standardowe podniesione do kwadratu. Kurtoza określa jak wiele obserwacji jest zbliżona do wartości średniej oraz czy większość z uzyskanych wyników ma wartość podobną do średniej. W badanym przypadku kurtoza wynosi 6,53 i jest większa od 0, co oznacza że jest więcej wartości odstających niż w rozkładzie normalnym oraz rozkład roku rejestracji jest wysmukły w porównaniu do rozkładu normalnego. Skośność wynosi -0,93 i świadczy to o lewostronnej asymetrii rozkładu. Zakres, czyli różnica między najwyższą a najniższą wartością, wynosi 70 lat. Najniższa wartość roku rejestracji wynosi 1953. Najwyższa wartość roku rejestracji wynosi 2023. Łączna suma wartości wynosi 7351165 lat. W badaniu uwzględniono 3654 rekordy. Współczynnik zmienności wynosi 0,25%. Informuje nas o małej zmienności i oznacza, że jest jednorodność badanej cechy.

1.4. Statystyka opisowa zmiennej Liczba poprzednich właścicieli

Tabela 4 Miary statystyczne zmiennej Liczba poprzednich właścicieli

Miary statystyczne	Liczba poprzednich właścicieli
Średnia	2,880131363
Błąd standardowy	0,020242521
Mediana	3
Tryb	3
Odchylenie standardowe	1,223626516
Wariancja próbki	1,49726185
Kurtoza	2,359422217
Skośność	0,88431653
Zakres	8
Minimum	1
Maksimum	9
Suma	10524
Licznik	3654
Współczynnik zmienności	0,42485094

Źródło: wynik funkcji Statystyka opisowa w programie MS Excel



Rysunek 2 Rozkład ilościowy samochodów z daną liczbą poprzednich właścicieli

Źródło: opracowanie własne

Z wykresu można wywnioskować, że kupując używany samochód będziemy jego 4 właścicielem. Liczba poprzednich właścicieli od 6 do 9 jest bardzo mało spotykana.

Liczba poprzednich właścicieli – ilość poprzednich właścicieli figurujących w dokumentach. Jest to zmienna ilościowa.

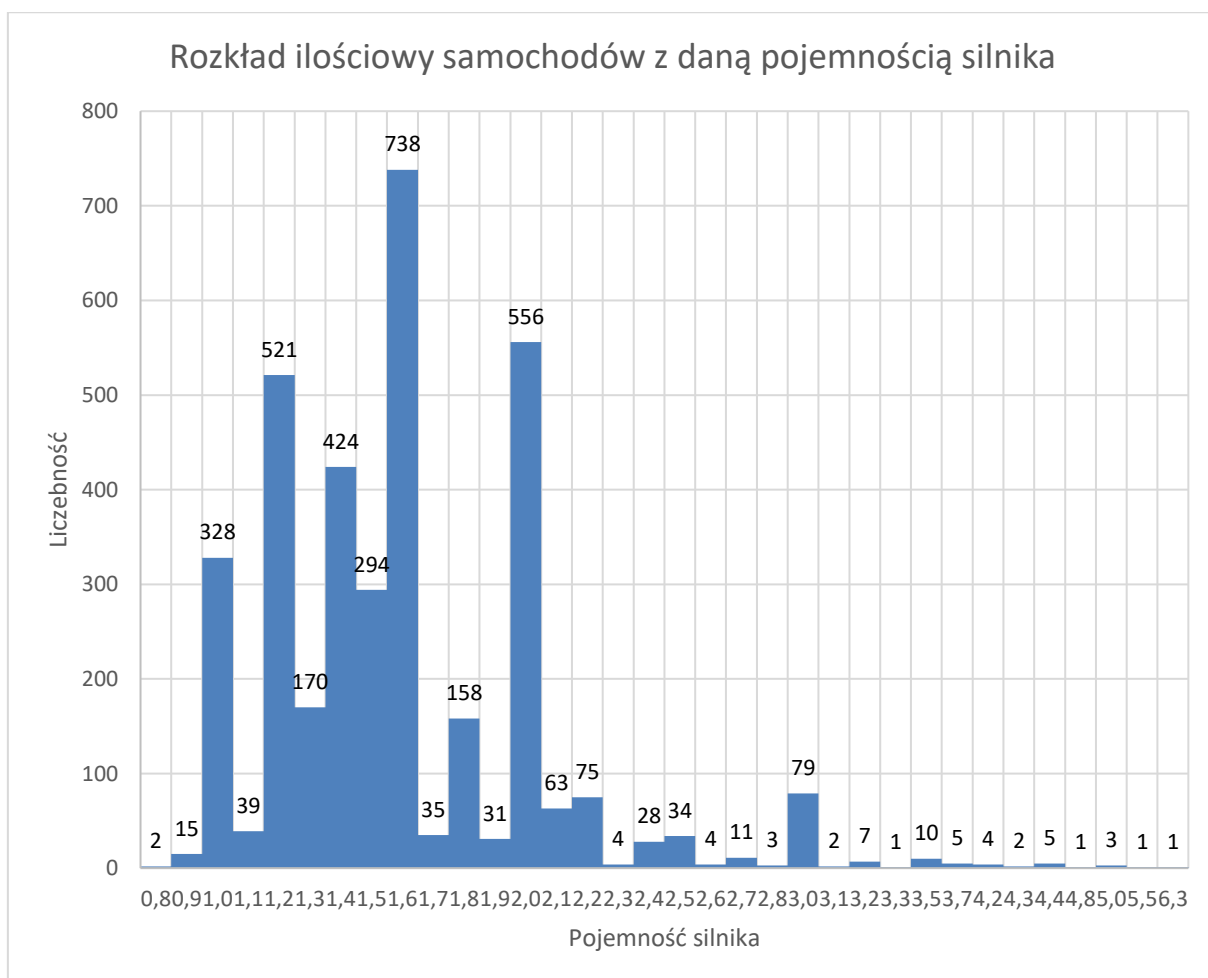
Średnia liczba poprzednich właścicieli dla auta wynosi w zaokrągleniu 3 osoby. Błąd standardowy wynosi 0,02 co oznacza, że gdyby przeprowadzane były próby losowe z tej samej populacji i obliczane by były wartości średniej dla każdej próby, to średnie z tych prób mogłyby się różnić od rzeczywistej średniej o 0,02. Mediana, czyli wartość środkowa, wynosi 3. Tryb zwraca najczęściej występującą wartość. Najczęściej występującym liczbą poprzednich właścicieli w zbiorze danych są 3 osoby. Odchylenie standardowe wynosi 1,23, oznacza to, że większość liczby poprzednich właścicieli w badanym zbiorze danych różni się od średniej liczby poprzednich właścicieli o 1,23. Wartość wariancji jest to odchylenie standardowe podniesione do kwadratu. Kurtoza określa jak wiele obserwacji jest zbliżona do wartości średniej oraz czy większość z uzyskanych wyników ma wartość podobną do średniej. W badanym przypadku kurtoza wynosi 2,36 i jest większa od 0, co oznacza że jest więcej wartości odstających niż w rozkładzie normalnym oraz rozkład liczby poprzednich właścicieli jest wysmukły w porównaniu do rozkładu normalnego. Skośność wynosi 0,88 i świadczy to o prawostronnej asymetrii rozkładu. Zakres, czyli różnica między najwyższą, a najniższą wartością przebiegu, wynosi 8 osób. Najniższa wartość liczby poprzednich właścicieli wynosi 1. Najwyższa liczba poprzednich właścicieli wynosi 9. Łączna suma liczby poprzednich właścicieli wynosi 10524 osoby. W badaniu uwzględniono 3654 rekordy. Współczynnik zmienności wynosi 42%. Informuje nas o przeciętnej zmienności i oznacza, że jest przeciętne zróżnicowanie cechy.

1.5. Statystyka opisowa zmiennej Pojemność silnika

Tabela 5 Miary statystyczne zmiennej Pojemność silnika

Miary statystyczne	Pojemność silnika
Średnia	1,606732348
Błąd standardowy	0,008056636
Mediana	1,6
Tryb	1,6
Odchylenie standardowe	0,487010126
Wariancja próbki	0,237178862
Kurtoza	9,411947811
Skośność	2,111250228
Zakres	5,5
Minimum	0,8
Maksimum	6,3
Suma	5871
Licznik	3654
Współczynnik zmienności	0,303105944

Źródło: Wynik funkcji Statystyka opisowa w programie MS Excel



Rysunek 3 Rozkład ilościowy samochodów z daną pojemnością silnika

Źródło: opracowanie własne

Z wykresu można wywnioskować, że kupujący kierują się aspektami ekonomicznymi w wyborze samochodu. Mniejszy litraż silnika skutkuje mniejszym poborem paliwa. Kupujący wybierają najczęściej samochody z silnikiem do dwóch litrów. Silniki z wysokim litrażem są bardzo rzadko wybierane przez kupujących.

Pojemność silnika – pojemność silnika to suma różnic pomiędzy maksymalną, a minimalną objętością każdego cylindra. Wartość ta jest wyrażana w cm sześciennych lub mniej precyzyjnie w litrach. W naszych danych pojemność silnika jest wyrażona w litrach. Jest to zmienna ilościowa.

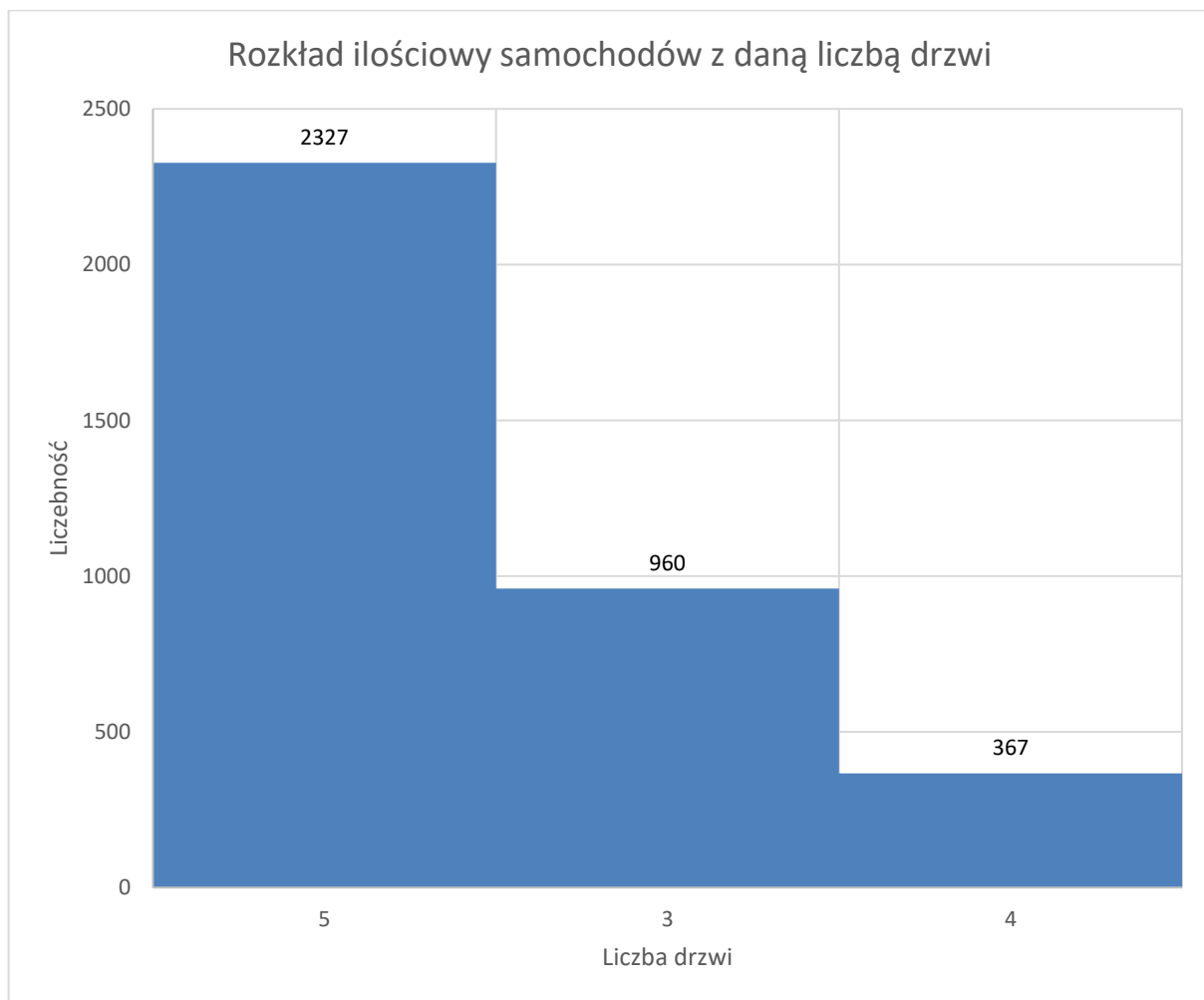
Średnie pojemności silników samochodów wynoszą 1,61 litra. Błąd standardowy wynosi 0,008 co oznacza, że gdyby przeprowadzane były próby losowe z tej samej populacji i obliczane by były wartości średniej dla każdej próby, to średnie z tych prób mogłyby się różnić od rzeczywistej średniej o 0,008. Mediana, czyli wartość środkowa, wynosi 1,6. Tryb zwraca najczęściej występującą wartość. Najczęściej występującą pojemnością silnika w zbiorze danych jest 1,6 litra. Odchylenie standardowe wynosi 0,49 litra, oznacza to, że większość pojemności silników w badanym zbiorze danych różni się od średniej pojemności silnika o 0,49 litra. Wartość wariancji jest to odchylenie standardowe podniesione do kwadratu. Kurtóza określa jak wiele obserwacji jest zbliżona do wartości średniej oraz czy większość z uzyskanych wyników ma wartość podobną do średniej. W badanym przypadku kurtóza wynosi 9,41 i jest większa od 0, co oznacza że jest więcej wartości odstających niż w rozkładzie normalnym oraz rozkład pojemności silnika jest wysmukły w porównaniu do rozkładu normalnego. Skośność wynosi 2,11 i świadczy to o prawostronnej asymetrii rozkładu. Zakres, czyli różnica między najwyższą, a najniższą pojemności silnika wynosi 5,5 litra. Najmniejsza pojemność silnika wynosi 0,8 litra. Największa pojemność silnika wynosi 6,3 litra. Łączna suma pojemności silników wynosi 5871 litrów. W badaniu uwzględniono 3654 rekordy. Współczynnik zmienności wynosi 30%. Informuje nas o przeciętnej zmienności i oznacza, że jest przeciętne zróżnicowanie cechy.

1.6. Statystyka opisowa zmiennej Liczba drzwi

Tabela 6 Miary statystyczne zmiennej Liczba drzwi

Miary statystyczne	Liczba drzwi
Średnia	4,374110564
Błąd standardowy	0,014420111
Mediana	5
Tryb	5
Odchylenie standardowe	0,871671583
Wariancja próbki	0,759811349
Kurtoza	-1,2038594
Skośność	-0,802065111
Zakres	2
Minimum	3
Maksimum	5
Suma	15983
Licznik	3654
Współczynnik zmienności	0,199279733

Źródło: Wynik funkcji Statystyka opisowa w programie MS Excel



Rysunek 4 Rozkład ilościowy samochodów z daną liczbą drzwi

Źródło: opracowanie własne

Z wykresu można wywnioskować, że w przeważającej ilości samochody posiadają 5 drzwi. Najprawdopodobniej są to auta o typie nadwozia Sedan. Ten typ nadwozia charakteryzuje się 5 drzwiami. Jest to najczęściej wybierany typ nadwozia w samochodach.

Liczba drzwi – liczba drzwi, które zawierają szybę. Jest to zmienna ilościowa.

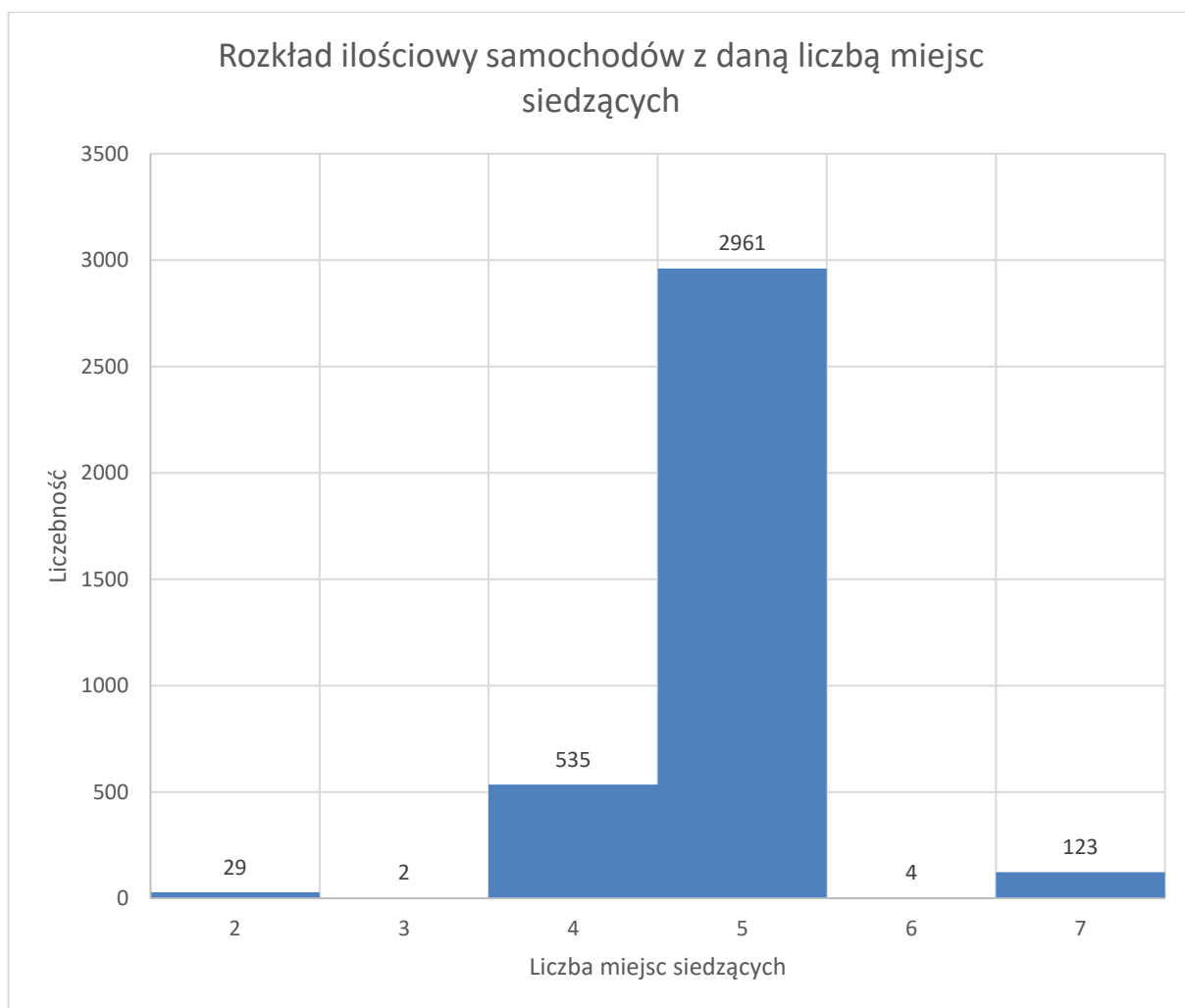
Średnia liczba drzwi w samochodzie wynosi w zaokrągleniu 4 drzwi. Błąd standardowy wynosi 0,001 co oznacza, że gdyby przeprowadzane były próby losowe z tej samej populacji i obliczane by były wartości średniej dla każdej próby, to średnie z tych prób mogłyby się różnić od rzeczywistej średniej o 0,001. Mediana, czyli wartość środkowa wynosi 5. Tryb zwraca najczęściej występującą wartość. Najczęściej występującą liczbą drzwi w zbiorze danych jest 5. Odchylenie standardowe wynosi 0,87, oznacza to, że większość liczby drzwi w badanym zbiorze danych różni się od średniej liczby drzwi o 0,87. Wartość wariancji jest to odchylenie standardowe podniesione do kwadratu. Kurtoza określa jak wiele obserwacji jest zbliżona do wartości średniej oraz czy większość z uzyskanych wyników ma wartość podobną do średniej. W badanym przypadku kurtoza wynosi -1,20 i jest mniejsza od 0, co oznacza że jest mniej dodatnich wartości odstających niż w rozkładzie normalnym oraz rozkład liczby drzwi w porównaniu do rozkładu normalnego jest spłaszczony. Skośność wynosi -0,80 i świadczy to o lewostronnej asymetrii rozkładu. Zakres, czyli różnica między najwyższą, a najniższą liczbą drzwi, wynosi 2. Najmniejsza liczba drzwi wynosi 3. Największa liczba drzwi wynosi 5. Łączna suma liczby drzwi wynosi 15983. W badaniu uwzględniono 3654 rekordy. Współczynnik zmienności wynosi 20%. Informuje nas o małej zmienności i oznacza, że jest małe zróżnicowanie cechy.

1.7. Statystyka opisowa zmiennej Liczba miejsc siedzących

Tabela 7 Miary statystyczne zmiennej Liczba miejsc siedzących

Miary statystyczne	Liczba miejsc siedzących
Średnia	4,89709907
Błąd standardowy	0,009720793
Mediana	5
Tryb	5
Odchylenie standardowe	0,587605638
Wariancja próbki	0,345280386
Kurtoza	8,099308591
Skośność	0,063923331
Zakres	5
Minimum	2
Maksimum	7
Suma	17894
Licznik	3654
Współczynnik zmienności	0,119990556

Źródło: Wynik funkcji Statystyka opisowa w programie MS Excel



Rysunek 5 Rozkład ilościowy samochodów z daną liczbą miejsc siedzących

Źródło: opracowanie własne

Z wykresu można wywnioskować, że przeważająca ilość samochodów posiada 5 miejsc siedzących. Wynika to z dostępności takich aut na rynku. Osoby konfigurujące nowy samochód najczęściej wybierają aspekty praktyczne samochodu. Do takich niewątpliwie należy ilość miejsc siedzących

Liczba miejsc siedzących – liczba miejsc siedzących udokumentowana w dowodzie rejestracyjnym. Jest to zmienna ilościowa.

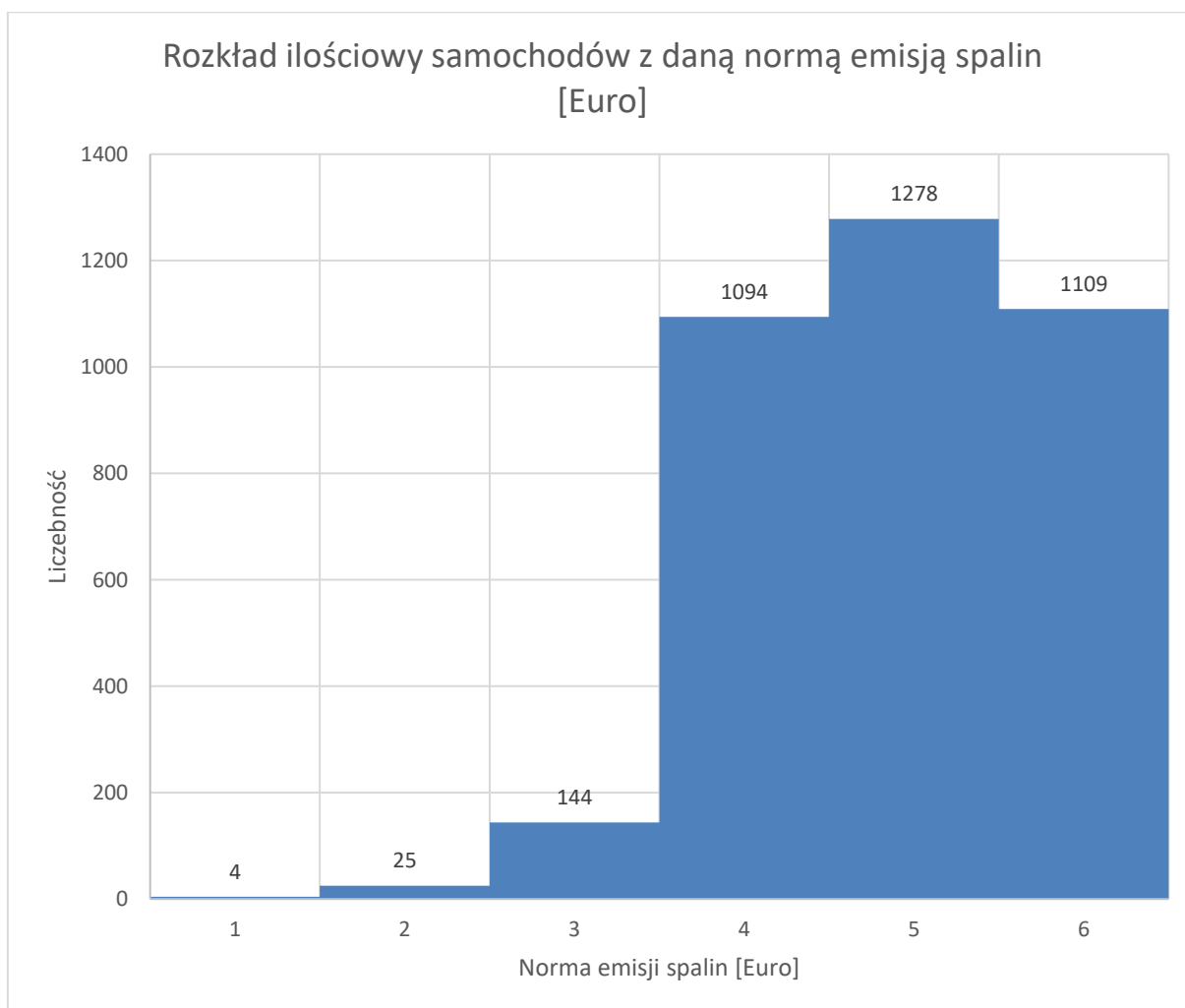
Średnia liczba miejsc siedzących w samochodzie wynosi w zaokrągleniu 5 miejsc. Błąd standardowy wynosi 0,01 co oznacza, że gdyby przeprowadzane były próby losowe z tej samej populacji i obliczane by były wartości średniej dla każdej próby, to średnie z tych prób mogłyby się różnić od rzeczywistej średniej o 0,01. Mediana, czyli wartość środkowa, wynosi 5. Tryb zwraca najczęściej występującą wartość. Najczęściej występującą liczbą miejsc siedzących w zbiorze danych jest 5. Odchylenie standardowe wynosi 0,59, oznacza to, że większość liczby miejsc siedzących w badanym zbiorze danych różni się od średniej liczby miejsc siedzących o 0,59. Wartość wariancji jest to odchylenie standardowe podniesione do kwadratu. Kurtoza określa jak wiele obserwacji jest zbliżona do wartości średniej oraz czy większość z uzyskanych wyników ma wartość podobną do średniej. W badanym przypadku kurtoza wynosi 8,10 i jest większa od 0 co oznacza że jest więcej wartości odstających niż w rozkładzie normalnym oraz rozkład liczby miejsc siedzących w porównaniu do rozkładu normalnego jest wysmukły. Skośność wynosi 0,06 i świadczy to o prawostronnej asymetrii rozkładu. Zakres, czyli różnica między najwyższą, a najniższą liczbą miejsc siedzących, wynosi 5. Najmniejsza liczba miejsc siedzących wynosi 2. Największa liczba miejsc siedzących wynosi 7. Łączna suma liczby miejsc siedzących wynosi 17894. W badaniu uwzględniono 3654 rekordy. Współczynnik zmienności wynosi 12%. Informuje nas o małej zmienności i oznacza, że jest małe zróżnicowanie cechy.

1.8. Statystyka opisowa zmiennej Norma emisji spalin [Euro]

Tabela 8 Miary statystyczne zmiennej Norma emisji spalin [Euro]

Miary statystyczne	Norma emisji spalin [Euro]
Średnia	4,900383142
Błąd standardowy	0,015070818
Mediana	5
Tryb	5
Odchylenie standardowe	0,911005717
Wariancja próbki	0,829931416
Kurtoza	-0,250258262
Skośność	-0,419628409
Zakres	5
Minimum	1
Maksimum	6
Suma	17906
Licznik	3654
Współczynnik zmienności	0,185904998

Źródło: Wynik funkcji Statystyka opisowa w programie MS Excel



Rysunek 6 Rozkład ilościowy samochodów z daną normą emisji spalin [Euro]

Źródło: opracowanie własne

Na wykresie można zobaczyć, że większość samochodów posiada Normę emisji spalin [Euro] od 4 do 6. Ilość samochodów z tą normą emisji spalin jest bardzo podobna. Waha się w okolicy 1100. Wynika to z tego, że na rynku aut używanych przeważają auta od rocznika 2005 do 2023. Im nowszy jest samochód tym norma emisji spalin [Euro] jest większa. W badanym zbiorze danych występuje mało samochodów, których rok pierwszej rejestracji jest mniejszy niż 2005.

Norma emisji spalin [Euro] – dopuszczalny poziom emisji spalin w pojeździe samochodowym. Jest to zmienna ilościowa.

Średnia wartość normy emisji spalin wynosi w zaokrągleniu 4,90. Błąd standardowy wynosi 0,02 co oznacza, że gdyby przeprowadzane były próby losowe z tej samej populacji i obliczane by były wartości średniej dla każdej próby, to średnie z tych prób mogłyby się różnić od rzeczywistej średniej o 0,02. Mediana, czyli wartość środkowa, wynosi 5. Tryb zwraca najczęściej występującą wartość. Najczęściej występującą wartością normy emisji spalin w zbiorze danych jest wartość 5. Odchylenie standardowe wynosi 0,91, oznacza to, że większość wartości normy emisji spalin w badanym zbiorze danych różni się od średniej wartości emisji spalin o 0,91. Wartość wariancji jest to odchylenie standardowe podniesione do kwadratu. Kurtoza określa jak wiele obserwacji jest zbliżona do wartości średniej oraz czy większość z uzyskanych wyników ma wartość podobną do średniej. W badanym przypadku kurtoza wynosi -0,25 i jest mniejsza od 0, co oznacza że jest mniej dodatnich wartości odstających niż w rozkładzie normalnym oraz rozkład normy emisji spalin w porównaniu do rozkładu normalnego jest spłaszczony. Skośność wynosi -0,42 i świadczy to o lewostronnej asymetrii rozkładu. Zakres, czyli różnica między najwyższą, a najniższą wartością normy emisji spalin wynosi 5. Najmniejsza wartość normy emisji spalin wynosi 1. Największa wartość normy emisji spalin wynosi 6. Łączna suma wartości normy emisji spalin wynosi 17906. W badaniu uwzględniono 3654 rekordy. Współczynnik zmienności wynosi 19%. Informuje nas o małej zmienności i oznacza, że jest małe zróżnicowanie cechy

2. Macierz korelacji Pearsona

Macierz korelacji Pearsona zawiera współczynniki korelacji między różnymi zmiennymi. Współczynniki korelacji mierzą siłę i kierunek zależności liniowej między dwiema zmiennymi. Macierz ta jest symetryczna, ponieważ korelacja między zmiennymi X i Y jest taka sama jak korelacja między Y i X. Macierz korelacji Pearsona zawiera jedynki na głównej przekątnej, ponieważ każda zmienna jest doskonale skorelowana sama ze sobą.

2.1. Macierz korelacji Pearsona ze zmiennymi z badania

	Cena[zt]	Przebieg[km]	Rok rejestracji	Liczba poprzednich właścicieli	Pojemność silnika	Liczba drzwi	Liczba miejsc siedzących	Norma emisji spalin [Euro]
Cena[zt]	1	-0,500343769	0,72426138	-0,401618848	-0,022834961	0,14138104	-0,03181723	0,693124668
Przebieg[km]	-0,500343769	1	-0,415179201	0,262505459	0,332234265	-0,03307965	0,094727588	-0,372787897
Rok rejestracji	0,72426138	-0,415179201	1	-0,393770327	-0,292343028	0,21737636	0,06630029	0,864733781
Liczba poprzednich właścicieli	-0,401618848	0,262505459	-0,393770327	1	0,146883272	-0,12476982	-0,032388813	-0,376618422
Pojemność silnika	-0,022834961	0,332234265	-0,292343028	0,146883272	1	-0,01167382	0,067278363	-0,262444264
Liczba drzwi	0,141381044	-0,033079651	0,217376364	-0,124769816	-0,011673817	1	0,434867975	0,186558509
Liczba miejsc siedzących	-0,03181723	0,094727588	0,06630029	-0,032388813	0,067278363	0,43486797	1	0,022778918
Norma emisji spalin [Euro]	0,693124668	-0,372787897	0,864733781	-0,376618422	-0,262444264	0,18655851	0,022778918	1

Rysunek 7 Macierz korelacji Pearsona

Źródło: Wynik funkcji Korelacja w programie MS Excel

Na podstawie współczynnika zmienności oraz metody Nowaka zostały odrzucone z modelu następujące zmienne: Przebieg [km], Rok rejestracji, Liczba poprzednich właścicieli, Pojemność silnika, Liczba drzwi. Do modelu wprowadzono dwie zmienne: Norma emisji spalin oraz Liczba miejsc siedzących. Wynika z tego, że zmienna objaśniana jest estymowana przez dwie zmienne objaśniające. Warto zwrócić uwagę na to, że dane nie są pozyskane z polskiego rynku samochodów używanych. Dane są z rynku angielskiego, więc zmienne determinujące cenę mogą się różnić od polskich.

3. Regresja

Do stworzenia modelu regresji użyto funkcji Regresja w zakładce Analiza danych w programie MS Excel. Wskazano zakres wejściowy Y, czyli Cenę [zł], a następnie zakres wejściowy X, czyli pozostałe, które wprowadzono do modelu. Są nimi Norma emisji spalin [Euro] oraz Liczba miejsc siedzących. Zaznaczono opcję Tytuły oraz przyjęto poziom ufności na poziomie 95%. Na samym końcu wskazano komórkę, w której miały się wygenerować wyniki.

3.1. Wynik funkcji Regresja

PODSUMOWANIE - WYJŚCIE								
Statystyki regresji								
Wielokrotność R	0,694758447							
R kwadrat	0,4826893							
Dopasowany R kwadrat	0,482405919							
Błąd standardowy	16689,02569							
Obserwacje	3654							
ANALIZA WARIANCJI								
	df	SS	MS	F	Istotność F			
Regresja	2	9,48833E+11	4,7442E+11	1703,327064	0,0000000			
Resztkowy	3651	1,01689E+12	278523579					
Razem	3653	1,96572E+12						
	Współczynniki	Błąd standardowy	t Stat	Wartość-p	Dolne 95%	Górne 95%	Dolne 95,0%	Górne 95,0%
Przecięcie	-47560,44294	2725,083903	-17,452836	1,40E-65	-52903,28	-42217,605	-52903,28047	-42217,60541
Liczba miejsc siedzących	-1880,339181	470,0379195	-4,0003989	6,44892E-05	-2801,902	-958,77628	-2801,902086	-958,7762766
Norma emisji spalin [Euro]	17676,89047	303,1780443	58,305312	0	17082,475	18271,3056	17082,47537	18271,30558

Rysunek 8 Funkcja Regresja

Źródło: Wynik funkcji Regresja w programie MS Excel

3.2. Statystyki regresji

Wielokrotność R - gdy spojrzymy na różnice w cenach samochodów w danych, model jest w stanie objąć (uwzględnić) 69% tych różnic, używając tylko dwóch zmiennych - "Normy emisji spalin" i "Liczby miejsc siedzących". To sugeruje, że te dwie zmienne są dość skuteczne w przewidywaniu cen samochodów w naszym modelu.

R kwadrat ($R \text{ kwadrat} \in [0;1]$) - wartość R kwadrat mówi o tym, jak dobrze zmienne objaśniające (w tym przypadku, "Norma emisji spalin" i "Liczba miejsc siedzących") wyjaśniają zmienność w cenie samochodu. Wartość 0,48 oznacza, że około 48% różnic w cenach samochodów można wyjaśnić przy użyciu tych dwóch zmiennych. 0,48 oznacza, że występuje niezadowalające dopasowanie prostej regresji do danych.

Przedziały dopasowania dla R^2 :

0.0 – 0.5 – dopasowanie niezadowalające

0.5 – 0.6 – dopasowanie słabe

0.6 – 0.8 – dopasowanie zadowalające

0.8 – 0.9 – dopasowanie dobre

0.9 – 1.0 – dopasowanie bardzo dobre

Dopasowany R kwadrat - wartość dopasowanego R kwadrat jest bardzo zbliżona do R kwadrat. To oznacza, że dodanie "Normy emisji spalin" i "Liczby miejsc siedzących" do modelu wniosło wartość i poprawiło zdolność do przewidywania cen samochodów. Wartość ta sugeruje, że te dwie zmienne są uzasadnione do uwzględnienia w modelu, mając na uwadze liczbę zmiennych i obserwacji.

Błąd standardowy - miara tego, jak bardzo rzeczywiste wartości danych różnią się od przewidywanych przez model. Im niższa wartość błędu standardowego, tym bardziej dane skupiają się wokół prognozowanej wartości.

Obserwacje – liczba obserwacji wynosi 3654. Jest to liczba przypadków, na podstawie których zbudowano model.

Istotność F mówi o istotności całego modelu. Jeżeli istotność $F < 0,05$ to oznacza że cały model jest statystycznie istotny. W tym przypadku można założyć, że model jest statystycznie istotny, ponieważ jego wartość jest bardzo mała i w tabeli wyświetla się jako 0.

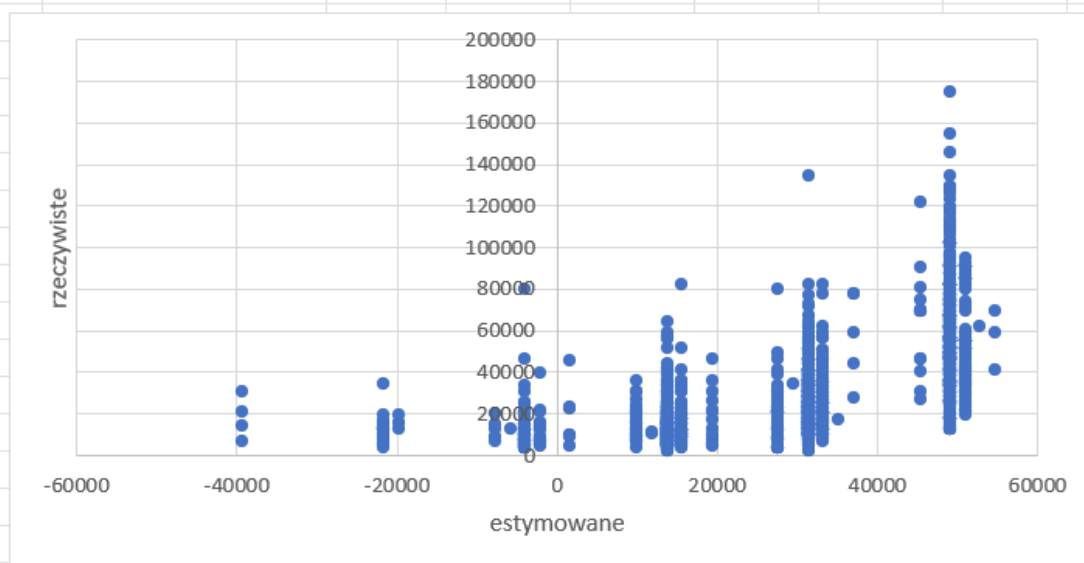
Wartość - p mówi o istotności poszczególnych parametrów. Jeśli wartość - $p < 0,05$, to parametr jest statystycznie istotny. W tym przypadku zmienne: Liczba miejsc siedzących oraz Norma emisji spalin i ich Wartości - p są $< 0,05$, więc te zmienne są statystycznie istotne.

3.3. Podsumowanie wyników regresji wielorakiej

Model regresji wielorakiej ma postać: $y = -47460,44294 - 1880,339181 \cdot x_7 + 17676,89047 \cdot x_8$, gdzie x_7 to Liczba miejsc siedzących, a x_8 to Norma emisji spalin [Euro]. Jeśli obie zmienne objaśniające są równe 0 to cena samochodu jest równa - 47460,44294 zł. Wraz ze wzrostem Liczby miejsc siedzących o 1 wartość, to cena samochodu spadnie o 1880,339181 zł. Może mieć to zastosowanie w rzeczywistości, ponieważ samochody z większą ilością miejsc siedzących są mniej rozchwytywane na rynku aut używanych. Druga zmienna objaśniająca to Norma emisji spalin [Euro]. Wraz ze wzrostem Normy emisji spalin [Euro] o 1 wartość, to cena samochodu rośnie o 17676,89047 zł. Jest to bardzo prawdopodobny scenariusz, ponieważ im wyższa Norma emisji spalin [Euro] tym nowszy jest samochód, a co za tym idzie jest droższy.

5.

Test normalności dla rezyduów



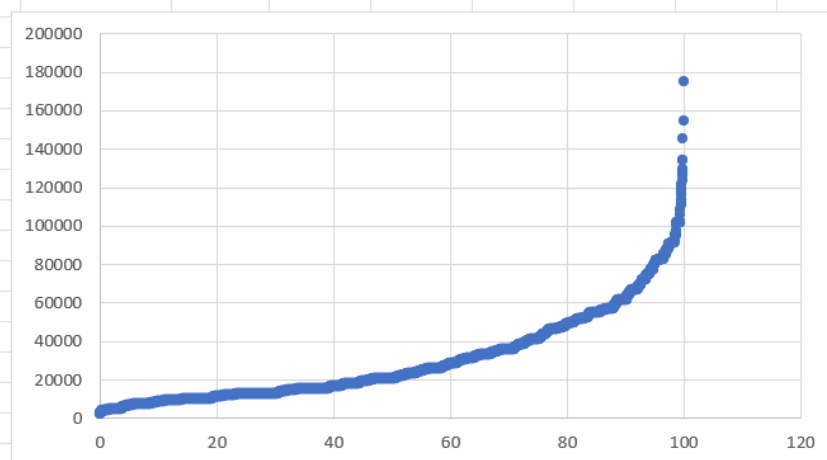
Wniosek: Brak rozkładu normalnego

Rysunek 13 Test normalności dla rezyduów

Źródło: opracowanie własne

5.

Test normalności dla rezyduów



Wniosek: Brak rozkładu normalnego, ponieważ krzywa nie jest znacząco prosta i nie jest równoległa do osi OX.

Rysunek 14 Test normalności dla rezyduów

Źródło: opracowanie własne

6.	Normalność						
	Test normalności dla rezyduów						
	H0: rezydua podlegają rozkładowi normalnemu						
	H1: rezydua nie podlegają rozkładowi normalnemu						
Wniosek:	H1: rezydua nie podlegają rozkładowi normalnemu						
	SKŁADNIKI RESZTOWE - WYJŚCIE						
	Obserwacja	Przewidywane Cena[zt]	Składniki resztowe				
	1	49099,20399	-13495,20399	odchylenie standardowe	16684,45649		
	2	13745,42305	-6031,223046	srednia	-9,41624E-10		
	3	31422,31352	-26525,47352	min	-36204,36399		
	4	15625,76223	-3267,562227	max	125824,796		
	5	31422,31352	-26262,31352	rozstep	162029,16		
	6	15625,76223	-11497,76223	liczba przedzialow	60,44832504	60	
	7	13745,42305	-9627,743046	rozpietosc	2700,486		

Rysunek 15 Test Normalność
Źródło: opracowanie własne

początek przedziału	koniec przedziału	liczebność	liczebność skumulowana	częstość względna skumulowana (dystrybuanta empiryczna (F(x))	dystrybuanta rozkładu normalnego F(u)
-36204,36399	-33503,87799	14	14	0,003831418	0,0223169033044
-33503,87799	-30803,39199	24	38	0,010399562	0,0324292000424
-30803,39199	-28102,90599	14	52	0,01423098	0,0460544671921
-28102,90599	-25402,41999	15	67	0,01833607	0,0639394092011
-25402,41999	-22701,93399	69	136	0,037219485	0,0868099983931
-22701,93399	-20001,44799	128	264	0,072249589	0,1153014444719

Rysunek 16 Test Normalność
Źródło: opracowanie własne

roznice	modul liczby				
D=max(F(x)-F(u))		Lambda = D*pierwiastek(n)	max modul liczby=	0,108149	lambda= 6,537419
-0,0184854856799	0,01848549				
-0,0220296379187	0,02202964	H0 odrzucamy			
-0,0318234874439	0,03182349				
-0,0456033391409	0,04560334	H1: rezydua nie podlegają rozkładowi normalnemu			
-0,0495905128978	0,04959051				
-0,0430518549809	0,04305185				

Rysunek 17 Test Normalność
Źródło: opracowanie własne

7. Homoskedastyczność

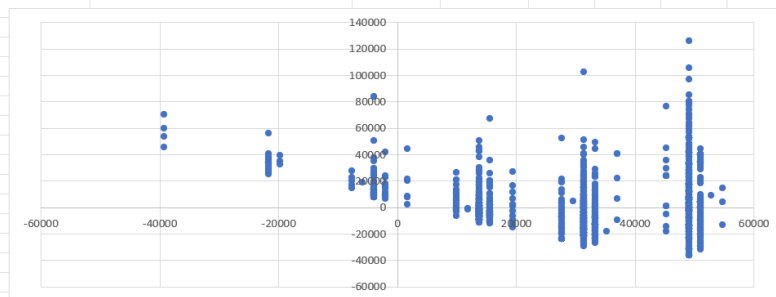
SKŁADNIKI RESZTOWE - WYJŚCIE

Observacja	Przewidywane Cena[zt]	Składniki resztowe
1	49099,20399	-13495,20399
2	13745,42305	-6031,223046
3	31422,31352	-26525,47352
4	15625,76223	-3267,562227
5	31422,31352	-26262,31352
6	15625,76223	-11497,76223
7	13745,42305	-9627,743046
8	13745,42305	-3451,223046
9	13745,42305	-9875,423046
10	13745,42305	-7042,583046
11	13745,42305	-6031,223046
12	31422,31352	-16200,31352
13	49099,20399	-13495,20399
14	49099,20399	53068,79601
15	15625,76223	-9949,762227
16	13745,42305	-7063,223046
17	13745,42305	-8590,583046
18	13745,42305	-7094,183046
19	13745,42305	-7553,423046
20	31422,31352	-21365,47352
21	13745,42305	-7570,223046

Test heteroskedastyczności

H0: heteroskedastyczność wariancji nie występuje (wariancja reszt jest jednorodna)

H1: heteroskedastyczność wariancji występuje (wariancja reszt jest niejednorodna)



Ponieważ punkty układają się haotycznie, heteroskedastyczność nie występuje (wariancja reszt jest jednorodna/stala)

Rysunek 18 Test Homoskedastyczność

Źródło: opracowanie własne

1 podpróba								
Nazwa	Cena[zt]	Przebieg[km]	Rok rejestracji	Liczba poprzednich właścicieli	Pojemność silnika	Liczba drzwi	Liczba miejsc siedzących	Norma emisji spalin [Euro]
SKODA Fabia	35604	113004,29	2016	3	1,4	5	5	6
Opel Corsa	7714,2	142621,85	2008	4	1,2	3	5	4
Hyundai i30	4896,84	220570	2011	3	1,4	5	5	5
MINI Hatch	12358,2	155736,91	2010	5	1,4	3	4	4
Opel Corsa	5160	136850	2013	3	1,3	5	5	5
Hyundai Coupe	4128	199955,56	2007	3	2,0	3	4	4
Ford Focus	4117,68	226364,39	2008	3	1,6	5	5	4
Opel Corsa	10294,2	144900	2009	3	1,2	3	5	4
Volvo 740	3870	362761,98	1989	3	2,3	5	5	4
Peugeot 207	6702,84	140070	2008	5	1,6	5	5	4

Rysunek 19 Test Homoskedastyczność

Źródło: opracowanie własne

2 podpróba								
Nazwa	Cena[zt]	Przebieg[km]	Rok rejestracji	Liczba poprzednich właścicieli	Pojemność silnika	Liczba drzwi	Liczba miejsc siedzących	Norma emisji spalin [Euro]
Alfa Romeo Mito	15454,2	132020	2013	3	1,4	3	4	5
Ford Mondeo	15454,2	181930	2011	3	1,6	5	5	5
Ford Focus	7734,84	182231,07	2008	2	1,6	5	5	4
Opel Astra	15454,2	98271,18	2007	3	1,6	3	5	4
Opel Astra	15454,2	98271,18	2007	3	1,6	3	5	4
Opel Astra GTC	26310,8	108031	2013	3	1,4	3	5	5
Mercedes-Benz C Cla:	12874,2	144095	2005	3	1,8	3	4	3
BMW 1 Series	56218,2	119140	2015	3	2,0	5	5	6
Volkswagen Passat	39990	147073,5	2014	2	2,0	5	5	5
BMW 4 SERIES	77400	175490	2015	3	3,0	5	5	6
Citroen C3	17802	99567,23	2010	2	1,4	5	5	5
Mercedes-Benz A Cla:	72214,2	170660	2019	2	1,3	5	5	6

Rysunek 20 Test Homoskedastyczność

Źródło: opracowanie własne

wariancja resztowa 1 podproba $Se^2=$	13300,1	-732,640042	-35563,56248
	352,045	579,971605	3275,65494
	0,43905	13262,09267	#N/D
	713,807	1824	#N/D
	2,5E+11	3,20811E+11	#N/D
		SSE	
wariancja resztowa 2 podproba $Se^2=$	20439,6	-2695,47009	-54727,85027
	484,503	681,6933897	4119,431982
	0,49492	18514,45965	#N/D
	893,64	1824	#N/D
	6,1E+11	6,2524E+11	#N/D
		SSE	

Rysunek 21 Test Homoskedastyczność

Źródło: opracowanie własne

		1827	n-liczba	
	$Se^2=SSE/(n-k-1)$		2	k-liczba zmiennych niezależnych
	Se^2 1 podproba=	175883102,1		
	Se^2 2 podproba=	342785216,1		
	max	min		
	342785216,1	175883102,1		
	F=	1,948938		
	f krytyczne	1,080093283		
	H_0 przyjmujemy			
Wniosek:	Występuje homoskedastyczność			

Rysunek 22 Test Homoskedastyczność

Źródło: opracowanie własne

8. Autokorelacja									
Test symetrii składników losowych				D1	D2	D			
H0: $\rho = 0$ (brak zależności)	Dla hipotezy alternatywnej $\rho > 0$	Dla hipotezy alternatywnej $\rho < 0$		1,92	1,92		1,722496526		
H0: $\rho > 0$	$d < d_1 - H_0$ odrzucaamy	$d \leq 4 - d_2 - H_0$ przyjmujemy	H0 odrzucaamy						
	$d_1 < d < d_2$ – nie podejmujemy żadnej decyzji	$4 - d_2 < d < 4 - d_1$ – nie podejmujemy żadnej decyzji							
	$d_2 \leq d - H_0$ przyjmujemy	$4 - d_1 \leq d - H_0$ odrzucaamy							
	<i>SS resztkowy</i>								
Test Durbin-Watsona	1,01689E+12								
Demp=suma(ei-ei-1)^2/suma(ei^2)	1,75159E+12		estymowane cena[zt]	Cena[zt]		składniki resztowe (cena-estymowane cena)			
				49099,20399	35604	-13495,20399			
				13745,42305	7714,2	-6031,223046	55711011,6		
				31422,31352	4896,84	-26525,47352	420014302		
				15625,76223	12358,2	-3267,562227	540930438		
de=	1,75159E+12	1,722496526		31422,31352	5160	-26262,31352	528758587		

Rysunek 23 Test Autokorelacja

Źródło: opracowanie własne

9.

Symetria

Test symetrii składników losowych

H0: $p_+=1/2$

p+frakcja dodatnia(reszty które mają wartosc dodatnia)

H1: $p_+\neq 1/2$

frakcja dodatnia nie jest rowna 1/2(liczba reszt o wartosci dodatniej musi być połowa)

Wniosek:

liczba >0 (m)

1508

m/n=

0,412698

te=

-10,71766034

wartosc krytyczna

1,9606136

Wniosek:

H0 odrzucamy

Frakcja dodatnia nie jest równa 1/2.

Obserwacja

Przewidywane Cena[zt]

kładniki resztowe

1

49099,20399

-13495,20399

2

13745,42305

-6031,223046

3

31422,31352

-26525,47352

4

15625,76223

-3267,562227

5

31422,31352

-26262,31352

6

15625,76223

-11497,76223

7

13745,42305

-9627,743046

Rysunek 24 Test Symetria

Źródło: opracowanie własne

10.	Losowość	Test serii		
	Test - liczby serii			
	H0: błąd modelu jest losowy			
	H1: błąd modelu nie jest losowy			
	Liczba serii=	832		
	Liczbwa wartości dodatnich=	1508		
	Liczbwa wartości ujemnych=	2146		
	tablica rozkładu serii			
	k1=alfa/2			
	k2=1-(alfa/2)			
	kE(-nieskonczoność;1713,88>U<3501,05;+nieskonczoność)			
	Wniosek: H0 odrzucamy. Błąd modelu nie jest losowy. Są podstawy do odrzucenia hipotezy o losowości reszt modelu.			
		poziom istotności	0,05	
		m =	1508	
		n =	3654	
		wartości krytyczne rozkładu normalnego	1,96	
		2m(n-m)/n + 1	1772,301587	
		pierwiastek(2m(n-m)(2m(n-m)-n)/n^2(n-1)	29,29847726	
		2m(n-m)/n	1771,301587	
		L1 =	1713,876572	
		L2 =	3501,049588	

Rysunek 25 Test Losowość

Źródło: opracowanie własne

Składniki resztowe	Liczba serii	Cena	składniki resztowe (cena-estymowane cena)	liczba serii
-13495,20399	1	2064	-29358,31352	1
-6031,223046	1	2322	-11423,42305	1
-26525,47352	1	2554,2	-28868,11352	1
-3267,562227	1	2574,84	-11170,58305	1
-26262,31352	1	2580	-11165,42305	1
-11497,76223	1	2580	-11165,42305	1
-9627,743046	1	3070,2	-10675,22305	1
-3451,223046	1	3302,4	-10443,02305	1
-9875,423046	1	3586,2	-10159,22305	1
-7042,583046	1	3606,84	7538,307428	2
-6031,223046	1	3612	-12013,76223	3
-16200,31352	1	3612	-12013,76223	3
-13495,20399	1	3612	-10133,42305	3
53068,79601	2	3612	25220,3579	4
-9949,762227	3	3612	-10133,42305	5
-7063,223046	3	3870	-9875,423046	5
-8590,583046	3	3870	7801,467428	6
-7094,183046	3	3870	-6114,744683	7
-7553,423046	3	3870	-9875,423046	7
-21365,47352	3	4076,4	-9669,023046	7
-7579,223046	3	4102,2	8033,667428	8
25710,47601	4	4102,2	8033,667428	8

Rysunek 26 Test Losowość

Źródło: opracowanie własne

11.	Koincydencja		
Pary zmiennych	sign(β_i)	sign(r_{ij})	Czy zachodzi koincydencja? (Tak/Nie)
Cena vs liczba miejsc siedzących	-	-0,03181723	Tak
Cena vs norma emisji spalin [euro]	+	0,693124668	Tak
Wniosek:	Zachodzi koincydencja dla obu par zmiennych		

Rysunek 27 Test Koincydencja

Źródło: opracowanie własne

5. Wnioski

Badanie polegało na określeniu jakie zmienne decydują o cenie auta. Na danych pobranych ze strony kaggle.com wykonano czynności, które miały odpowiedzieć na to pytanie. Wszystkie etapy badania zostały wykonane w programie Microsoft Excel. W pierwszej kolejności wyznaczono miary statystyczne wszystkich zmiennych. Kolejnym etapem było stworzenie histogramów. Po tych czynnościach przy pomocy metody Nowaka oraz współczynnika zmienności odpowiedziano na pytanie jakie zmienne determinują cenę samochodu. Następnym etapem było stworzenie wzoru, który objaśniał cenę auta. Do tego potrzebne było wygenerowanie wyników z funkcji Regresja w zakładce Analiza Danych. Dzięki temu można było wywnioskować, że jeśli obie zmienne objaśniające są równe 0 to cena samochodu jest równa -47460,44294 zł. Wraz ze wzrostem Liczby miejsc siedzących o 1 wartość, to cena samochodu spadnie o 1880,339181 zł. Wraz ze wzrostem Normy emisji spalin [Euro] o 1 wartość, to cena samochodu rośnie o 17676,89047 zł. Ostatnim etapem było wykonanie testów, które odpowiadały na pytanie czy nasze dane oraz poprzednie wyniki są dobre do analizowania. Już w pierwszym teście (Zestawienie sum kwadratów) wyszło, że nasz model należy uznać za błędny. Już w tym momencie należało zaprzestać wykonywania dalszych testów, lecz na potrzeby projektu wykonano je wszystkie. W drugim teście (Ocena jakości dopasowania modelu) wyszło, że model jest niezadowolający, ponieważ w 48% wyjaśnia kształtowanie się zmiennej objaśnianej. W trzecim teście (Test istotności modelu) okazało się, że przynajmniej jeden z parametrów modelu jest istotnie różny od zera. Dzięki czwartemu testowi (Wnioskowanie o istotności parametrów strukturalnych) można było się dowiedzieć, że wszystkie zmienne przyjęte do badania są statystycznie istotne. W piątym teście (Test normalności dla rezyduów) okazało się, że rezydua nie podlegają rozkładowi normalnemu. Była do graficzna interpretacja wyników przedstawionych na wykresie. W kolejnym teście (Normalność) potwierdził się wniosek jaki postawiono w poprzednim teście. W teście, który dowodził czy występuje heteroskedastyczność, okazało się, że heteroskedastyczność nie występuje, ponieważ punkty na wykresie układają się chaotycznie. Potwierdziły to wyniki obliczeń. W ósmym teście (Autokorelacja) wyszło, że wśród reszt nie występuje autokorelacja pierwszego rzędu. W dziewiątym teście (Symetria) okazało się, że frakcja dodatnia nie jest równa 1/2. W dziesiątym teście (Losowość) po wykonaniu niezbędnych obliczeń dowiedziano się, że błąd modelu jest losowy i są podstawy do odrzucenia hipotezy o losowości reszt modelu. W ostatnim teście (Koincydencja) okazało się, że dla par zmiennych Cena vs Liczba miejsc siedzących oraz Cena vs Norma emisji spalin [Euro] zachodzi koincydencja.