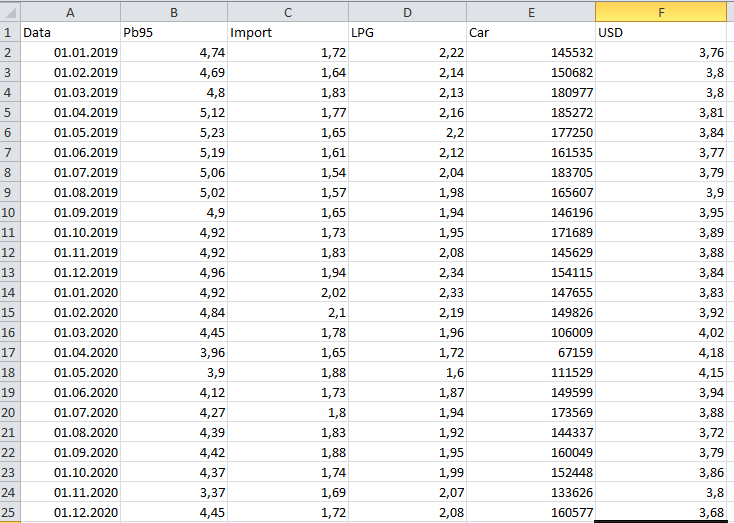
Sylwester Dylik

Nr indeksu: 145820

**Regresja liniowa**

Do wykonania regresji liniowej wykorzystałem język R (Rstudio) oraz własną bazę danych. W projekcie postanowiłem przedstawić regresję liniową ceny paliwa, która miała swoje wzloty jak i upadki w czasach pandemii. Ceny paliw zostały pobrane jako średnie miesięczne ze strony autocentrum.pl, ceny dolara z archiwum giełdy zaś liczbę zarejestrowanych pojazdów pobrałem z archiwum CEPiK.



Data – data określająca miesiąc pobrania danych

Pb95 – cena benzyny – zmienna objaśniana

Import – import ropy do polski w (mln ton) – zmienna objaśniająca

LPG – cena LPG – zmienna objaśniająca

Car – liczba nowo zarejestrowanych samochodów – zmienna objaśniająca

USD – kurs dolara w zł – zmienna objaśniająca

Arkusz kalkulacyjny z bazą potrzebnych nam danych zapisałem w formacie .xlsx

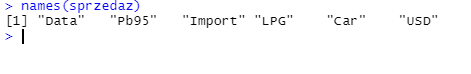
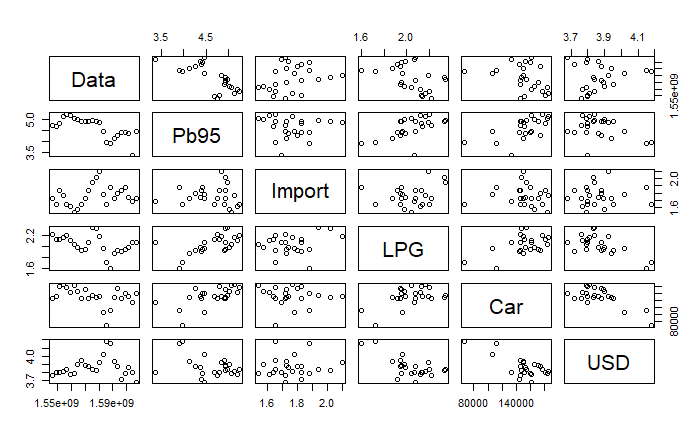
oraz umieściłem w folderze zawierającym projekt regresji liniowej.

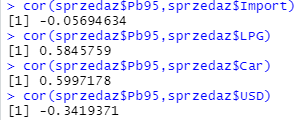
Na początku należy zainstalować bibliotekę pozwalającą nam na wczytanie naszej bazy .

Następnie wczytujemy naszą bibliotekę zaznaczając lokalizację danych tylko na arkuszu 1 oraz wyświetlamy załadowane dane.



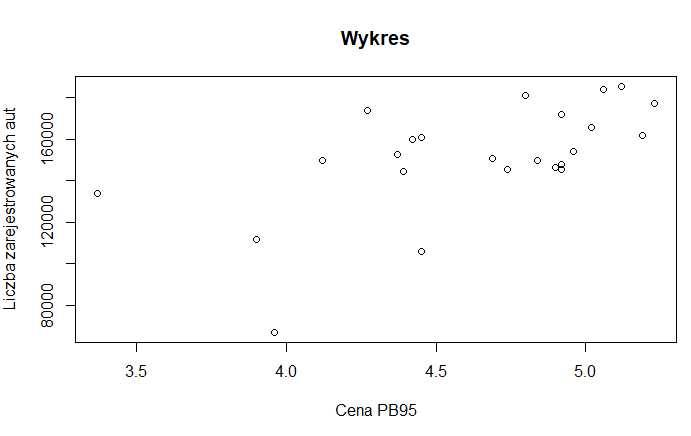
Sprawdzam poprawność załadowanych tabel czy poprawnie widzę nazwy kolumn w programie.

Sprawdzam współzależność danych od sibie: 



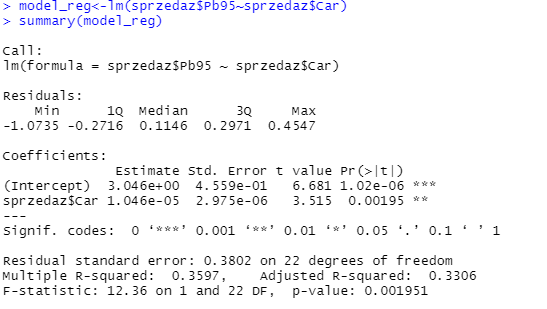
Możemy zauważyć że największą współzależność naszej zmiennej objaśnianej PB95 ma zmienna Car.

Wykres względem tych zmiennych:





Kolejnym krokiem naszej regresji liniowej jest wyznaczenie parametrów naszej linii jak i określenie wielkości błędu, w języku r możemy to bardzo prosto otrzymać:



Przy wyznaczaniu parametrów naszej regresji wkradł się błąd w zapisie dotyczącym wyświetlenie nam parametrów, niestety zjadłem jedną literę.

Opis parametrów:

1) Median - Mediana

2) Estimate – Odchylenie standardowe

3) Std. Error – Oszacowanie błędu

4) t value – sprawdza czy współczynnik jest różny od 0

5) Pr(>|t|) – badanie istotności modelu jako całości

6) Residual standard error – resztkowy błąd standardowy

7)Multiple R-squared – współczynnik

8)Adjusted R-squared – uwzględnienie liczby zmiennych w modelu

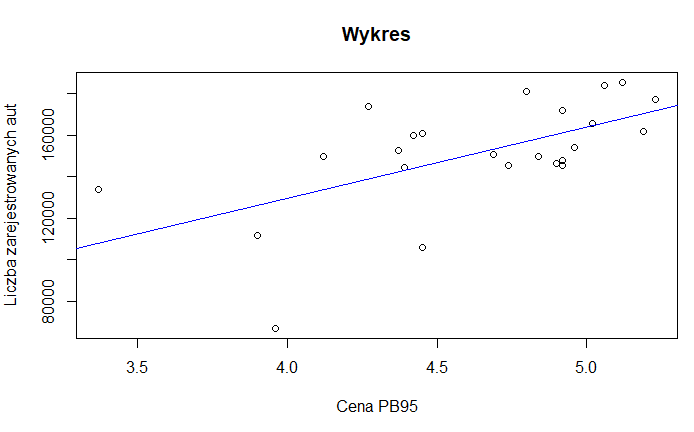
9)F-statistic – parametr mówiący nam czy regresja ma sens

Obliczamy przedział ufności :



Teraz rysujemy naszą regresję liniową:





Analiza tabeli odchyleń:

