**Proiect**

**Analiza datelor in R**

**Student: Miucin Sebastian**

In acest referat voi analiza datele din setul de date : mtcars.

Datele din acest set au fost extrasa dintr-o revista din 1974 numita Motor Trend US care descrie consumul de combustibil si alte 10 aspecte de design si performanta pentru 32 de automobile.

**Formatul setului de date:**

Un set de date cu 32 de observatii pe 11 variabile numerice. - “[,x]” – coloana

[,1]mpg – Mile / gallon (consumul)

[,2]cyl – numarul de cilindrii pe care ii are motorul

[,3]disp – marimea motorului (cc – centimetrii cubi)

[,4]hp – cai putere

[,5]drat – tractiune spate

[,6]wt – greutatea masinii ( in lb )

[,7]qsec – timpul scos pe ¼ mile

[,8]vs – alinierea cilindrilor ( 0 = in forma de V , 1 = in linie)

[,9]am – cutie de viteze (0 = automata , 1 = manuala)

[,10]gear – numarul de viteze

[,11]carb – numarul de carburatoare

**Statistica descriptiva**

**Summary**

Aplicand instructiunea “**summary”** pe variabile , putem afla diferiti indicatori numerici cum ar fi **:** valoarea cea mai **mica** si cea mai **mare, prima** si a 3-a **quartila** , **media** si **mediana** .

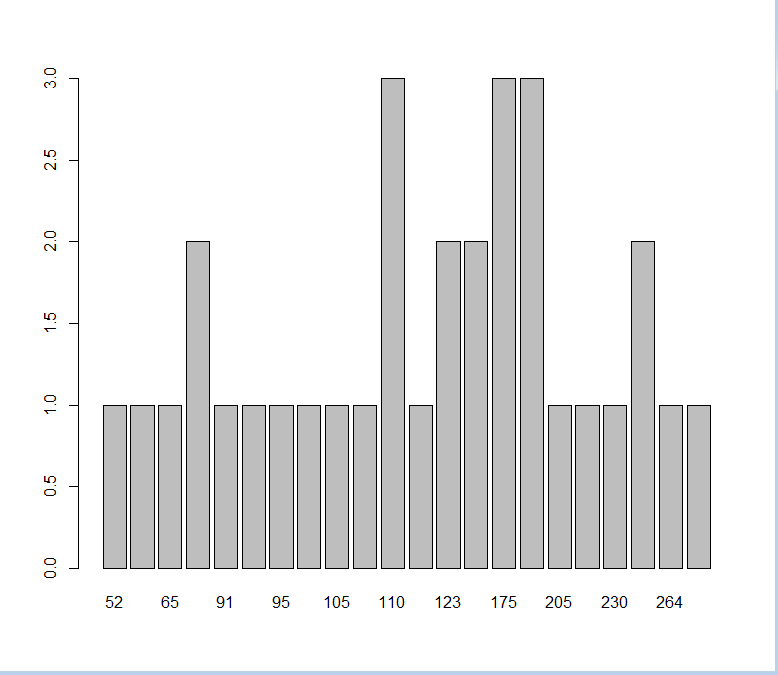
|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Variabila** | **Minim** | **1 quartila** | **Mediana** | **Media** | **3 quartila** | **max** |
| Hp | 52.0 | 96.5 | 123.0 | 146.7 | 180.0 | 335 |
| Mpg | 10.40 | 15.43 | 19.20 | 20.09 | 22.80 | 33.90 |
| Cyl | 4 | 4 | 6 | 6.18 | 8 | 8 |
| Ws | 1.51 | 2.58 | 3.32 | 3.21 | 3.61 | 5.42 |

**Dispersia de selectie + abatere standard:**

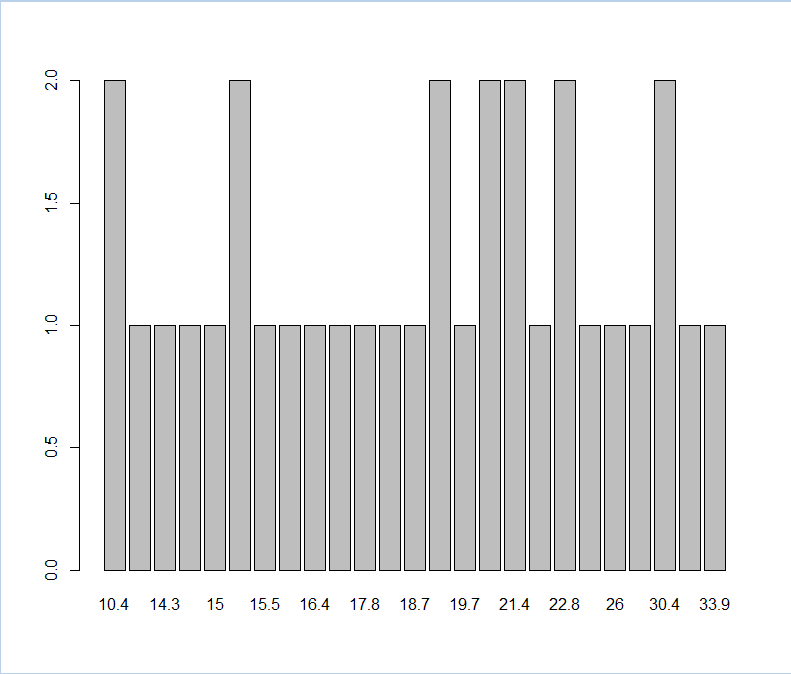
Aplicand instructiunea **“var”** pe variabile, putem afla **dispersia de selectie** , iar **“skewness”** ne va arata **abaterea standard.**

|  |  |  |
| --- | --- | --- |
| **Variabila** | **Dispersie** | **Abatere** |
| Hp | 4700 | 0.76 |
| Mpg | 36.32 | 0.64 |
| Cyl | 3.18 | -0.18 |
| Wt | 0.95 | 0.44 |

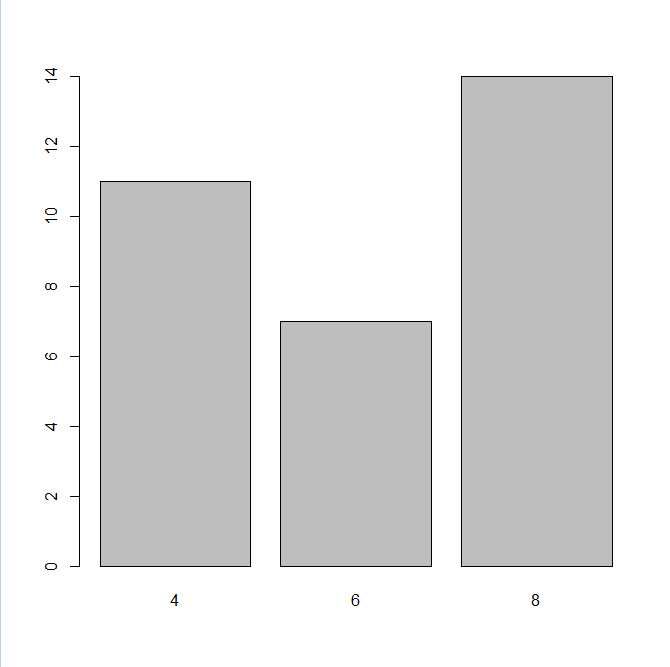
**Barplot-uri**

Utilizand instructiunea “barplot(table(x))” pe variabila “hp” putem vedea ca cele mai multe masini cu acelasi numar de cai putere sun t cele cu 110 , 175 si 180 de cai putere.

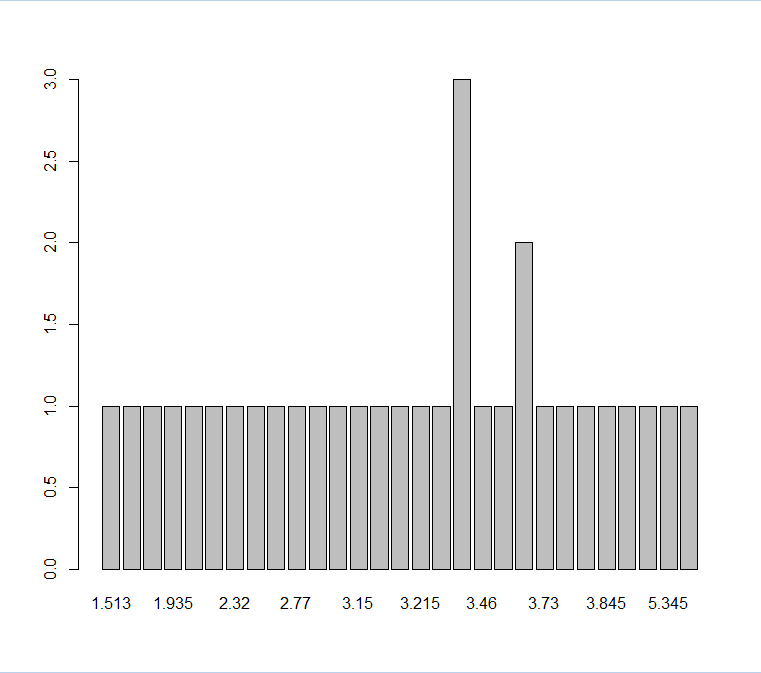
Utilizand instructiunea “barplot(table(x))” pe variabila “mpg” putem vedea ca masinile care au acelasi consum in comun sunt de cel mult doua.



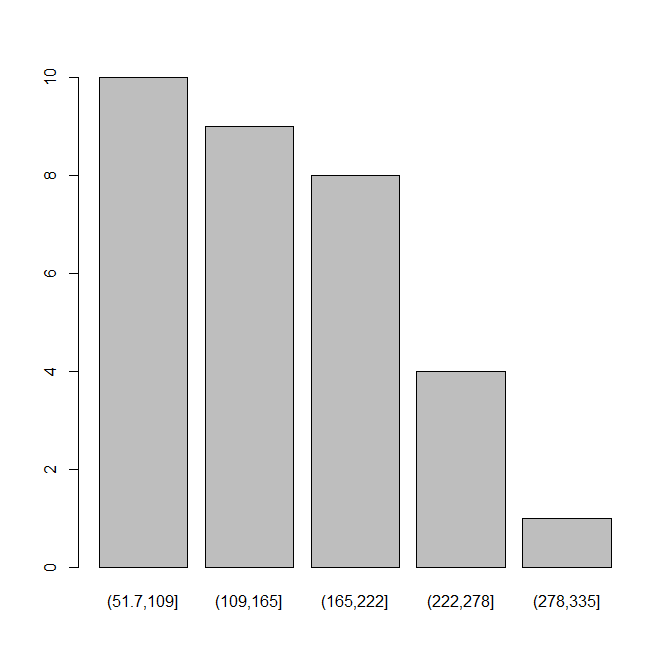
Utilizand instructiunea “barplot(table(x))” pe variabila “cyl” putem vedea ca mjoritatea masinilor din acest set de date au motor cu 8 cilindri.



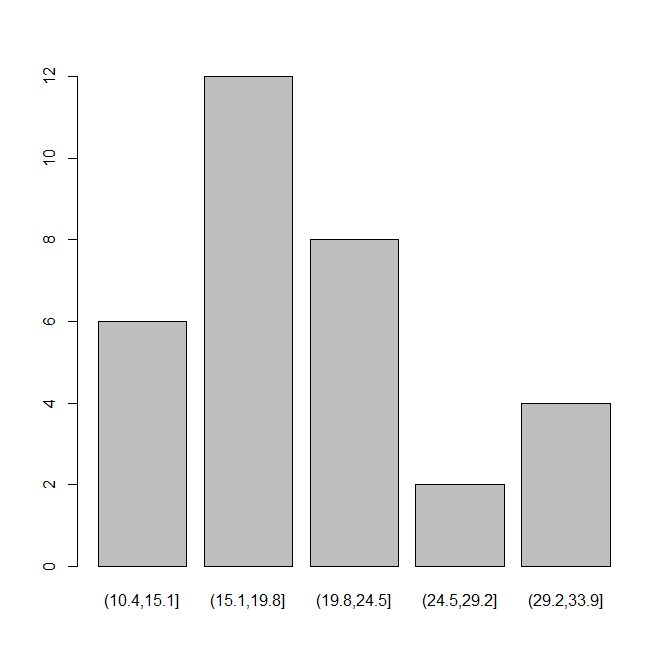
Utilizand instructiunea “barplot(table(x))” pe variabila “cyl” putem vedea ca cele mai multe masini care au aceeasi greutate sunt in numar de 3 (3.44)



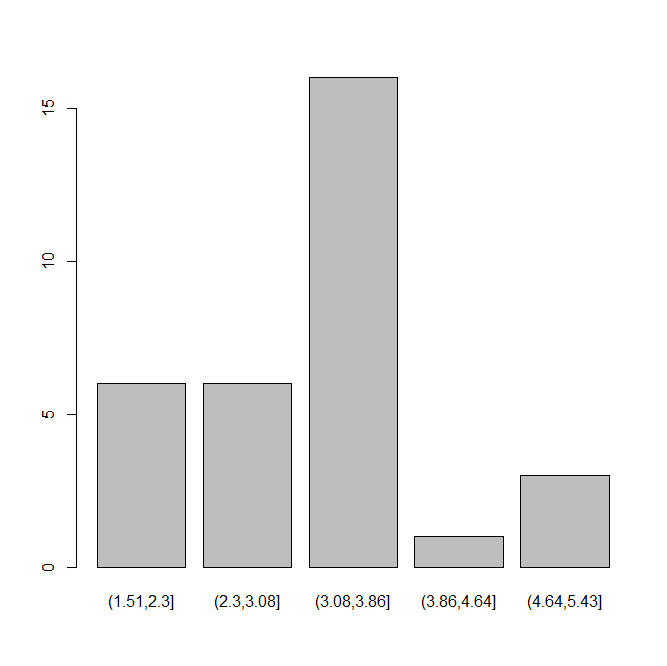
Utilizand instructiunea “barplot(table(cut(x,5)))” pe variabila “hp” putem vedea ca majoritatea masinilor au intre 51 si 109cp



Utilizand instructiunea “barplot(table(cut(x,5)))” pe variabila “mpg” putem vedea ca majoritatea masinilor merg intre 15.1 - 19.8 mile cu 1 galon de combustibil.



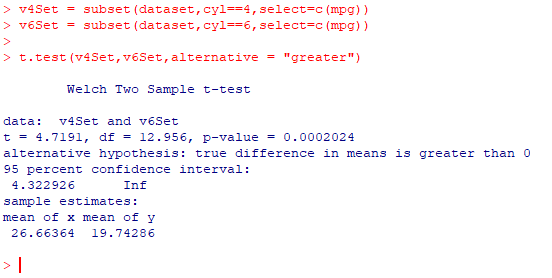
Utilizand instructiunea “barplot(table(cut(x,5)))” pe variabila “wt” putem vedea ca majoritatea masinilor cantaresc intre 3.08 si 3.86lb.



**Ipoteze**

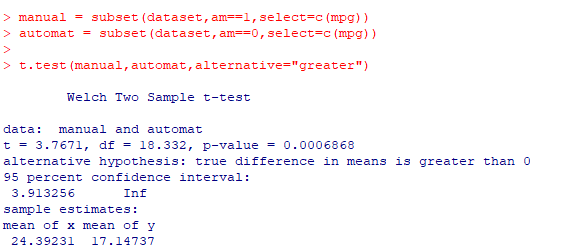
**Ipoteza 1** : Consumul masinilor cu 4 cilindri este mai redus decat a celor cu 6. **=> Fals**

Pentru a verifica aceasta ipoteza vom creea doua subseturi de date doar cu coloanele “cyl” si “mpg” apoi rulam testul.



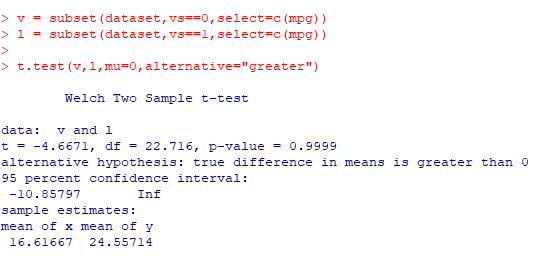
**Ipoteza 2** : Consumul masinilor este mai ridicat daca masina are transmisie manuala **=> Adevar**

Pentru a verifica aceasta ipoteza vom creea doua subseturi de date doar cu coloanele “am” si “mpg” apoi rulam testul.

****

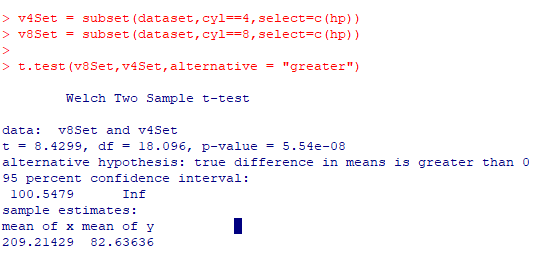
**Ipoteza 3** : Daca motorul este in forma de v consumul creste **=> Fals**

Pentru a verifica aceasta ipoteza vom creea doua subseturi de date doar cu coloanele “vs” si “mpg” apoi rulam testul.



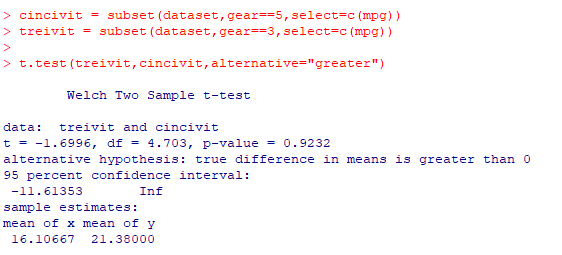
**Ipoteza 4** : Masinile cu mai multi cilindri au mai multi cai putere **=> Adevarat**

Pentru a verifica aceasta ipoteza vom creea doua subseturi de date doar cu coloanele “cyl” si “hp” apoi rulam testul.



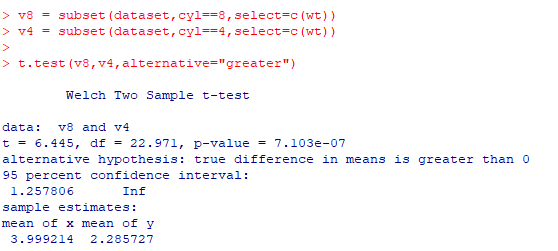
**Ipoteza 5** : Masinile cu 3 trepte consuma mai mult decat cele cu 5 **=> Fals**

Pentru a verifica aceasta ipoteza vom creea doua subseturi de date doar cu coloanele “gear” si “mpg” apoi rulam testul.



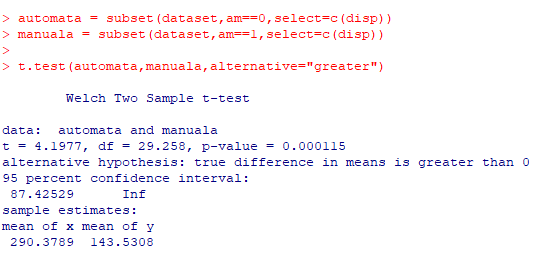
**Ipoteza 6** : masinile care au motoare cu mai multi cilindi cantaresc mai mult **=> Adevarat**

Pentru a verifica aceasta ipoteza vom creea doua subseturi de date doar cu coloanele “cyl” si “wt” apoi rulam testul.



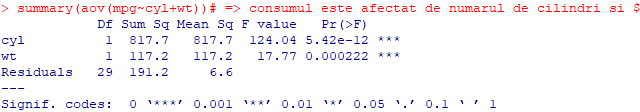
**Ipoteza 7** : Masinile cu cutie automata au motoare mai mari(cc) **=> Adevarat**

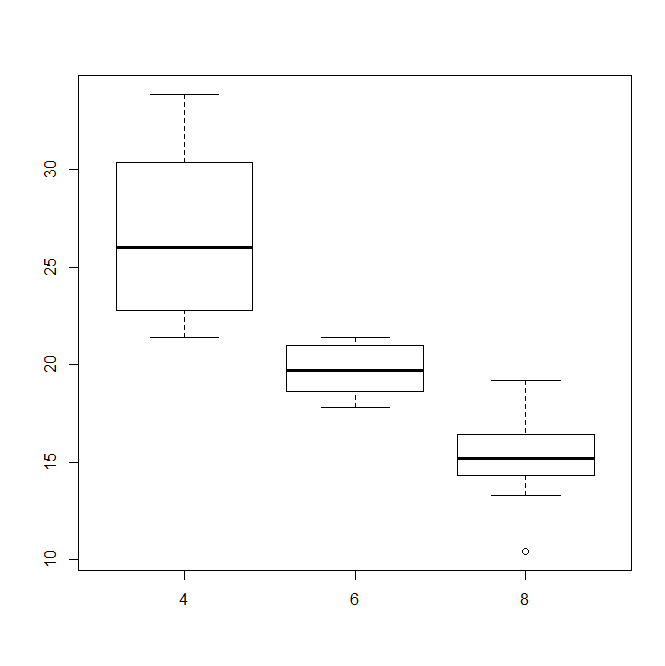
Pentru a verifica aceasta ipoteza vom creea doua subseturi de date doar cu coloanele “am” si “disp” apoi rulam testul.

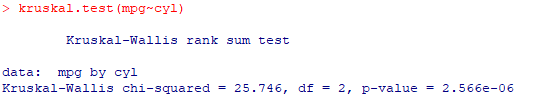


**Teste ANOVA/Kruskal**

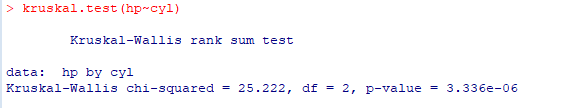
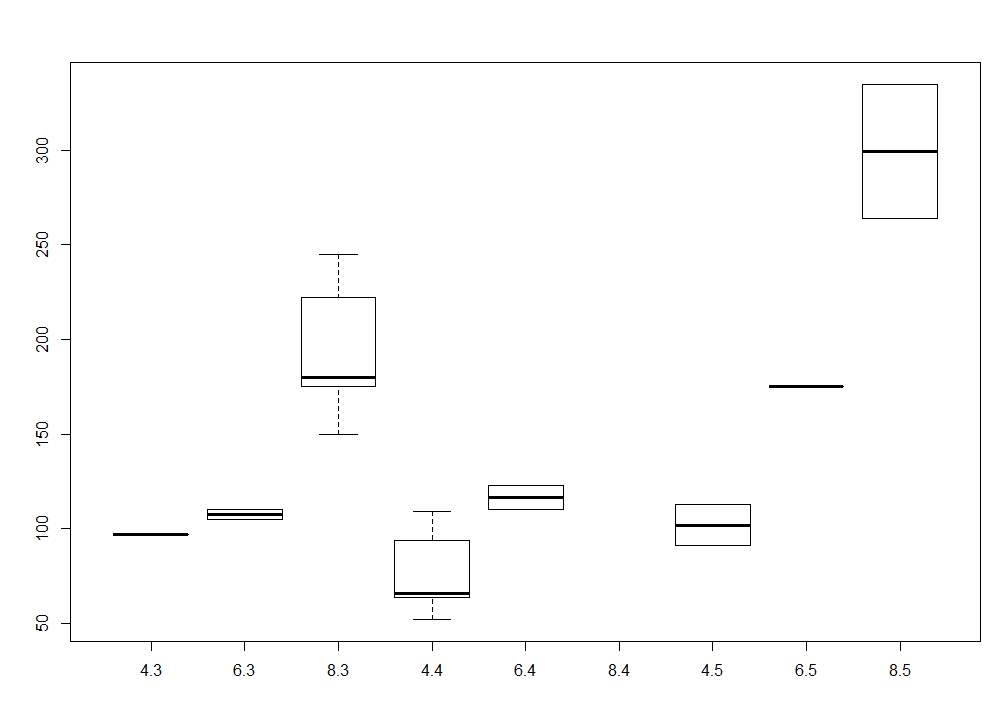
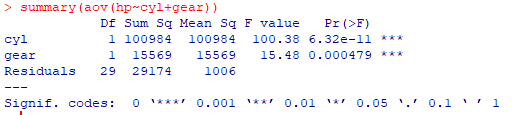
-Consumul este afectat in functie de numarul de cilindri si de greutate

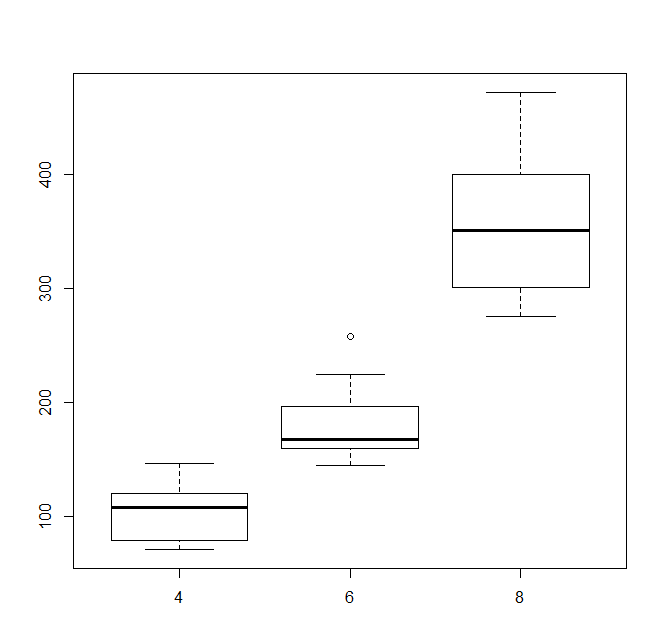
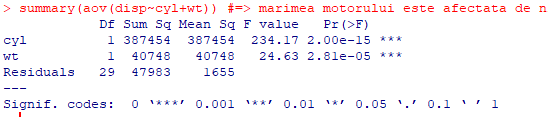


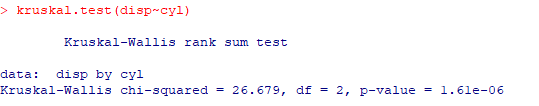




-Masina are mai multi cai putere in functie de numarul de cilindii si numarul de viteze

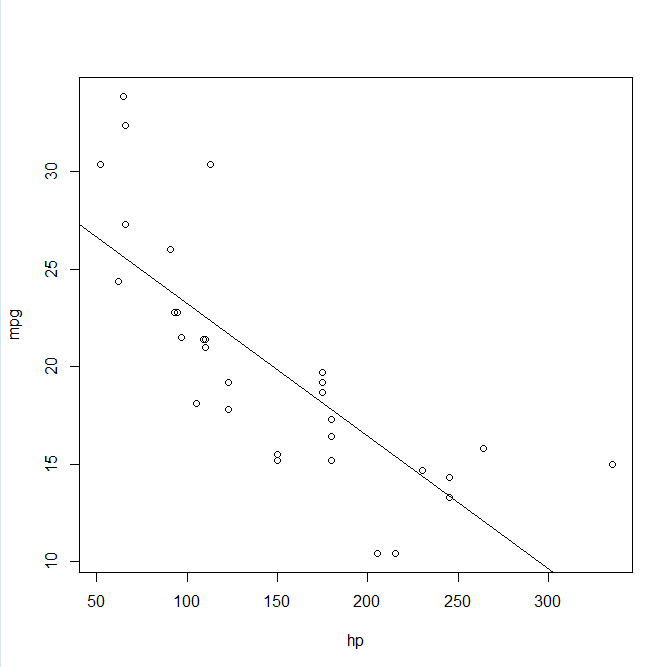
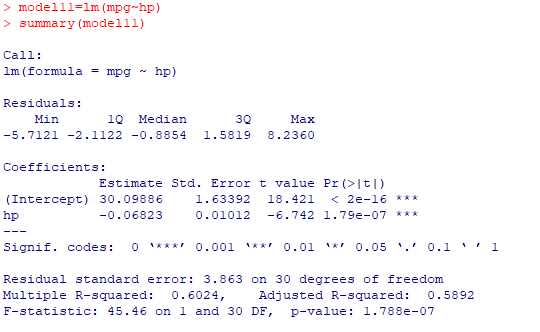


-Marimea motorului este afectata de numarul de cilindii si de greutatea masinii

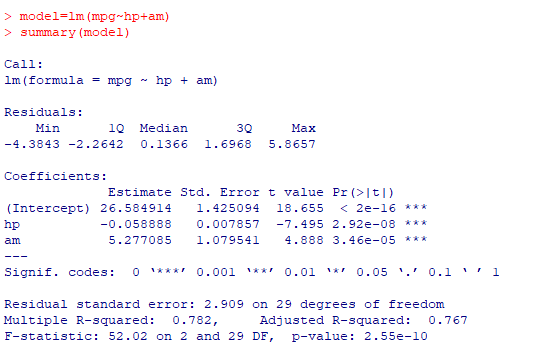


**Regresii liniare unifactoriale/multifactoriale**

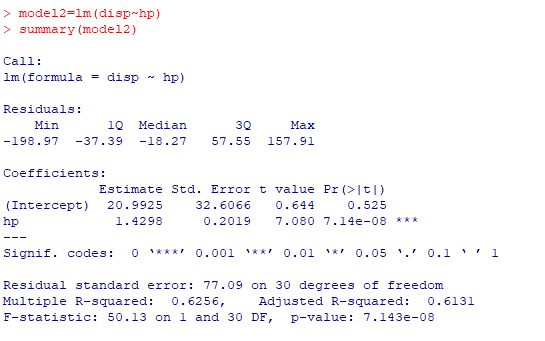
-Consumul creste in functie de numarul de cai putere

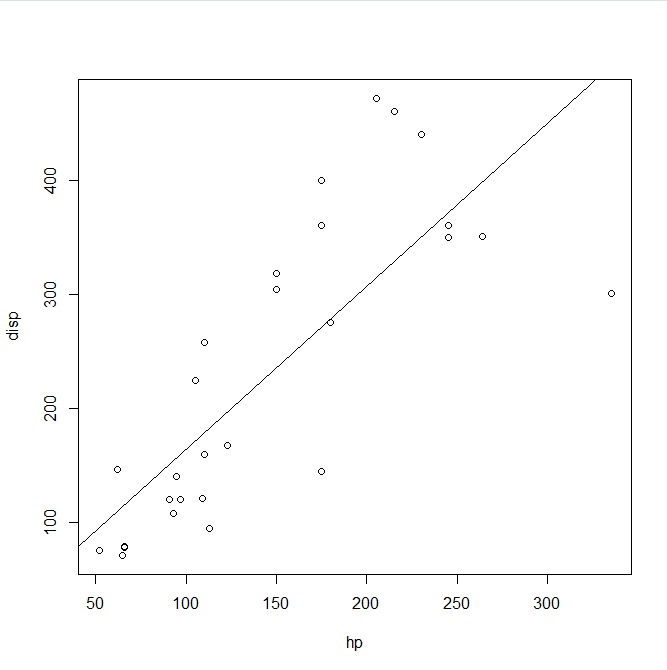


dar si in functie de tipul cutiei de viteze:

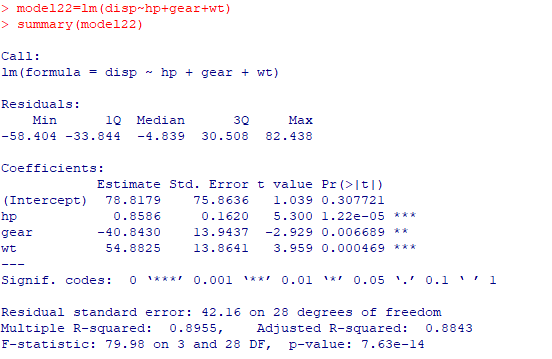


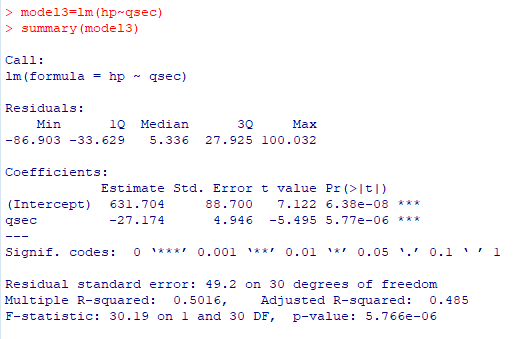
-Marimea motoarelor sunt afectate de numarul de cai putere

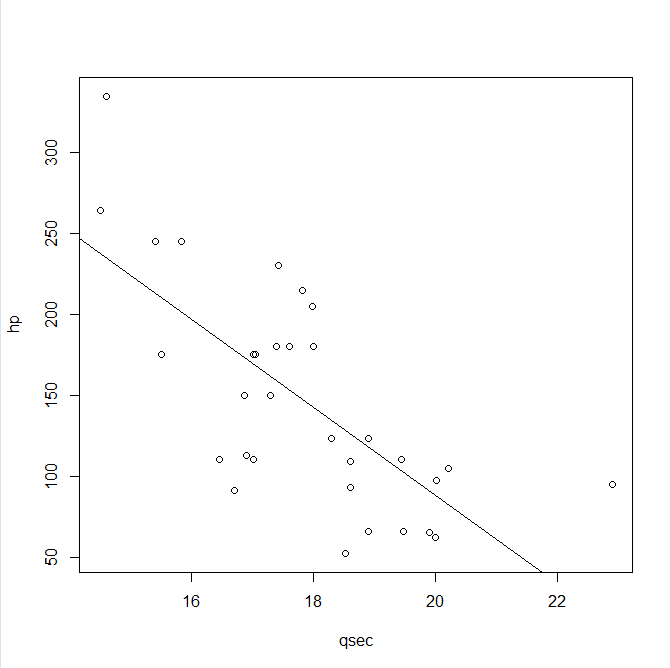




dar si de numarul de viteze si de greutatea masinii:



-Caii putere sunt influentati de timpul qsec si de capacitatea motorului:  




dar si de capacitatea motorului:

