Business Analytics
Assignment 3
Yash Verma
U101113FCS167

Q1. Boston Housing

Target variable = MEDV.

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
MEDV	506	5	50	22.52	9.209
Valid N (listwise)	506				

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.826[a]	.683	.674	5.258

Initial regression summary for the data with all predictors.

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 29.0946152  4.3504779   6.688 6.30e-11 ***
CRIM        -0.1126687  0.0336252  -3.351 0.000869 ***
ZN           0.0398325  0.0144821   2.750 0.006175 **
INDUS       -0.0574903  0.0598258  -0.961 0.337053
CHAS         2.7239398  0.8942639   3.046 0.002446 **
NOX         -1.5479430  0.7654187  -2.022 0.043692 *
RM           3.3773311  0.4082532   8.273 1.29e-15 ***
AGE          0.0002086  0.0139030   0.015 0.988038
DIS         -1.2587267  0.1966385  -6.401 3.66e-10 ***
RAD          0.2480576  0.0675867   3.670 0.000269 ***
TAX         -0.0110656  0.0039065  -2.833 0.004811 **
PTRATIO     -0.8425516  0.1292512  -6.519 1.79e-10 ***
B            0.0098812  0.0027452   3.599 0.000352 ***
LSTAT       -0.5906061  0.0503389 -11.733  < 2e-16 ***
```

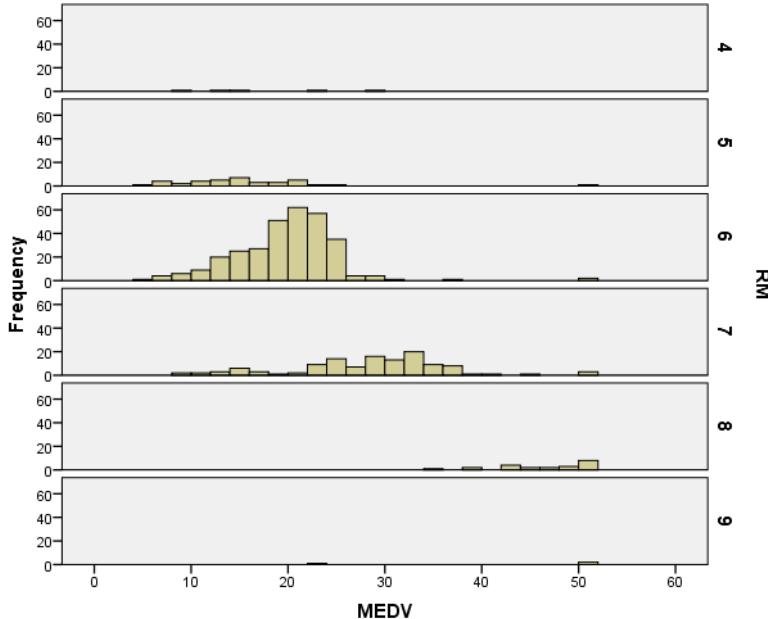Box plot with cook's distance. This gives us information of outlier values.
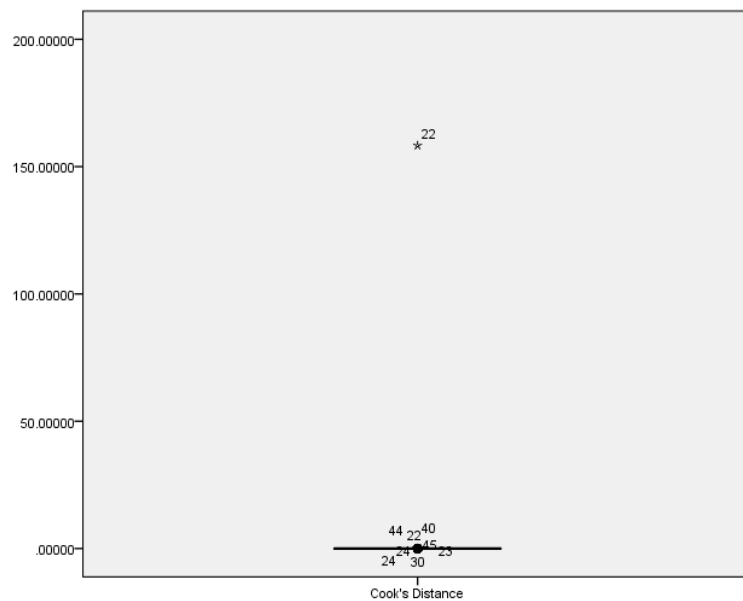
We use select cases and remove all the outlier values starting from the bottom to avoid the changing of numbering below.

A quick analysis using R shows the significant variables. The variables used are -
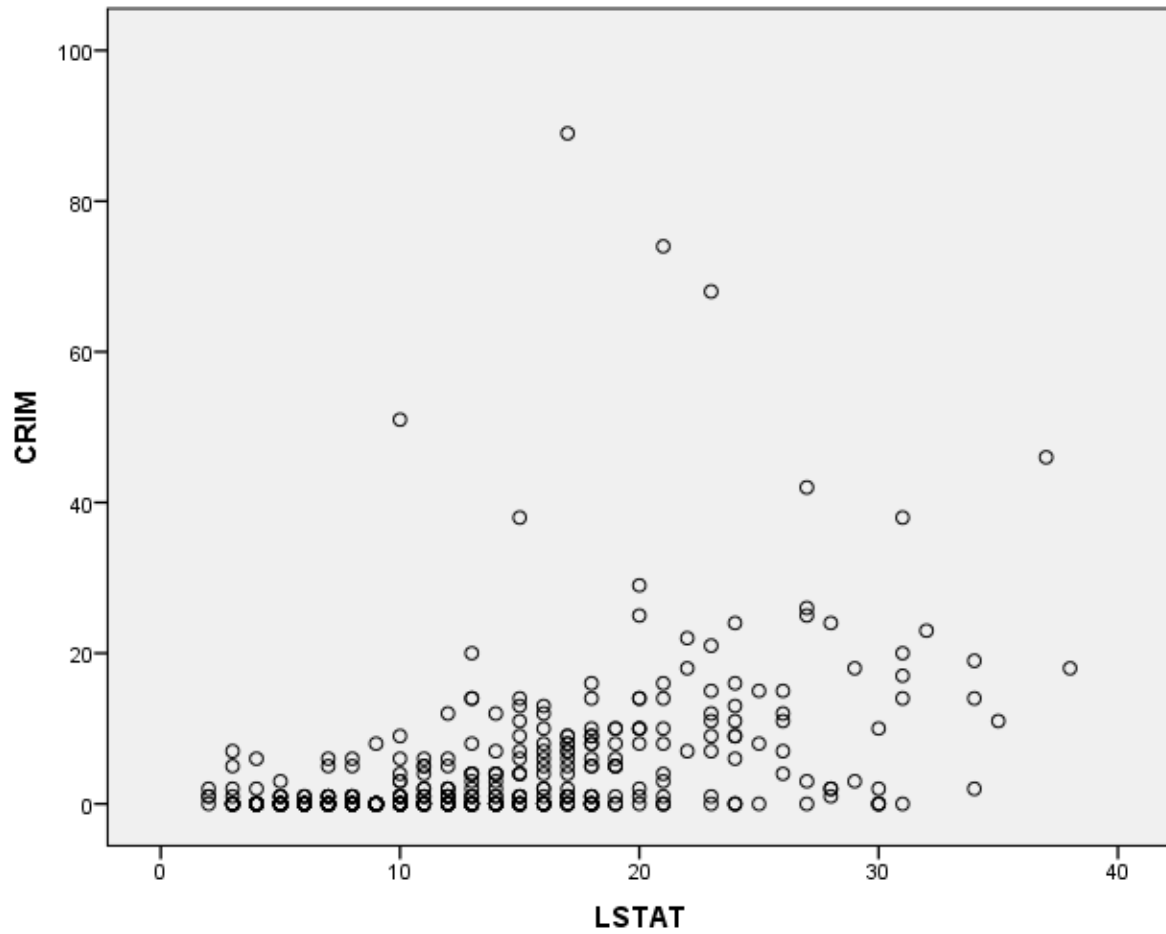
- RM
 The number of rooms in each house affect the price of the house greatly.

- LSTAT
  The % of lower status population in general leads to more crime and the graph below supports the claim. Therefore, if the LSTAT value is high, it affects the prices of the houses inversely.

- PTRATIO
  The pupil to teacher ration by town depicts the schools in the town. A higher ratio of pupil to teacher means more schools in the town, increase the cost of homes.
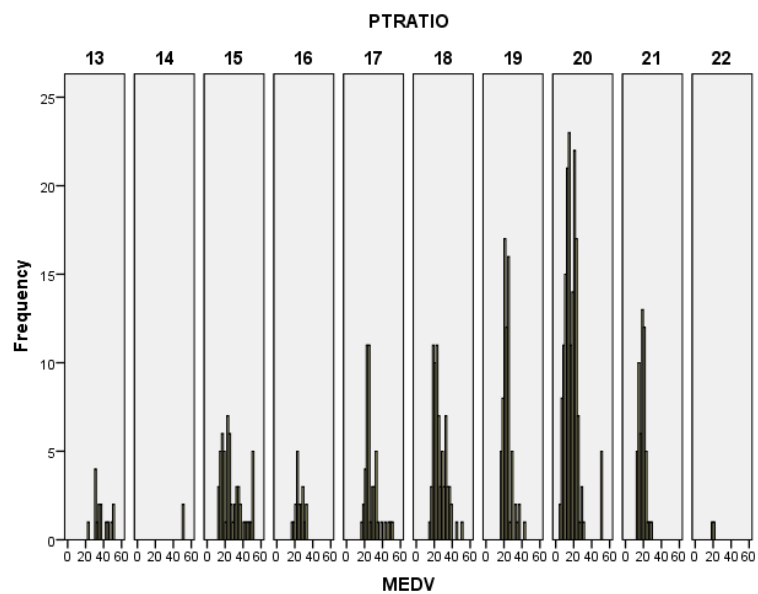
- RAD

  Easy access to highways or major roads imply general development of the area, the more an area is developed, the higher is the price of the house in the area.
- CRIM

  The crime in an area affects the value of homes. The higher the crime rate in a town/area, the lower the prices of the houses owned and vice-versa.
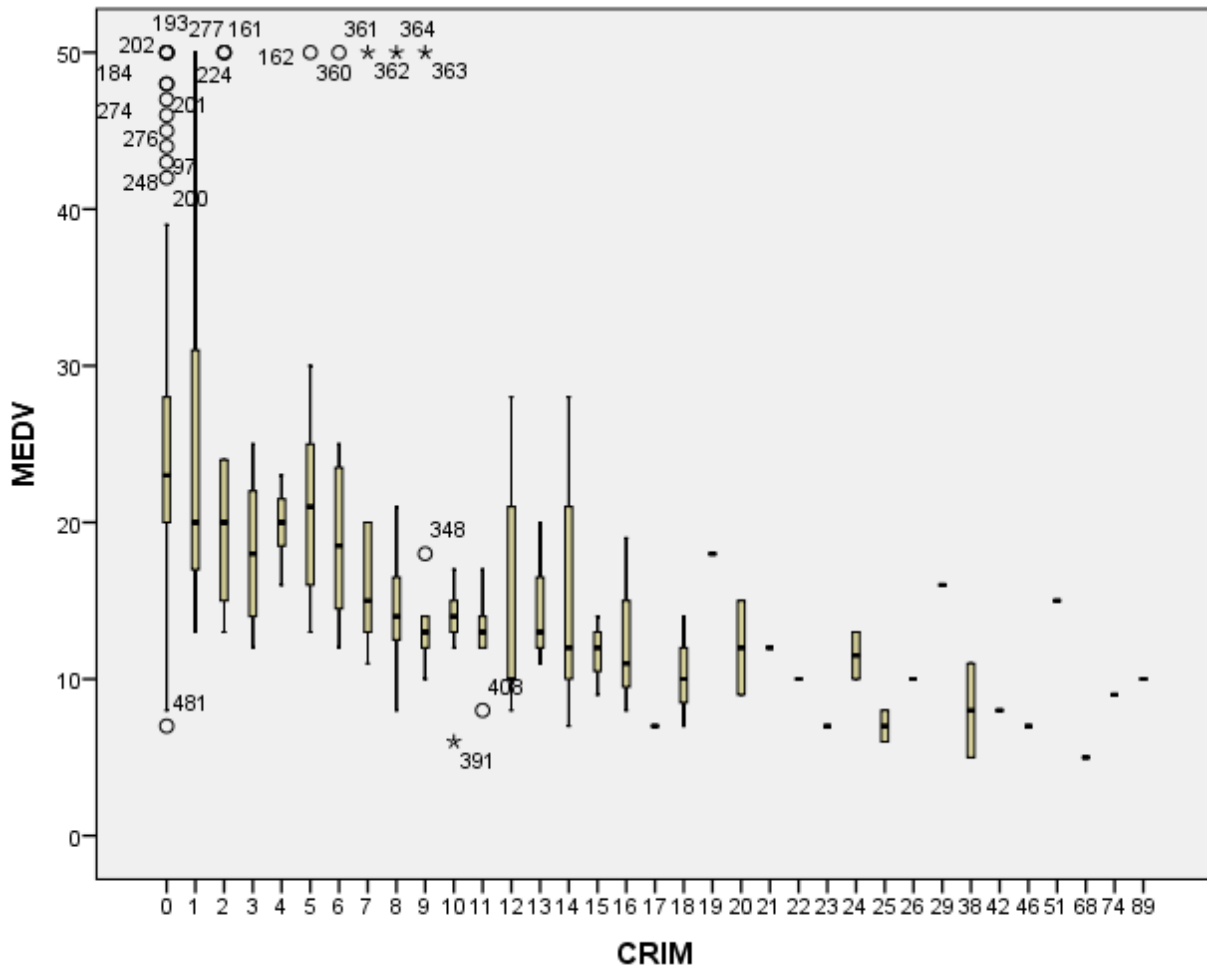
- TAX

  The full-value of property-tax rate affects the cost of the houses.
- CHAS

  Properties along the river side have a higher value tag because of the scenic view.
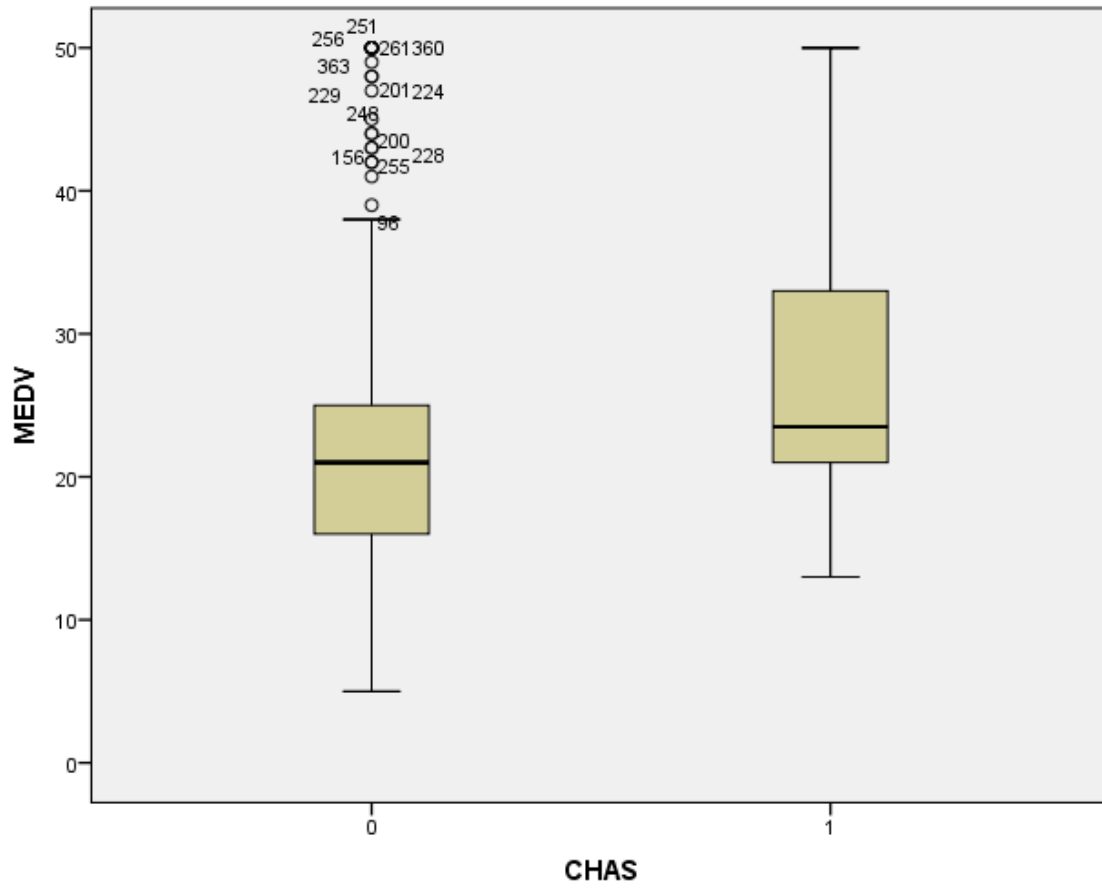
The importance of the variables are as follows -
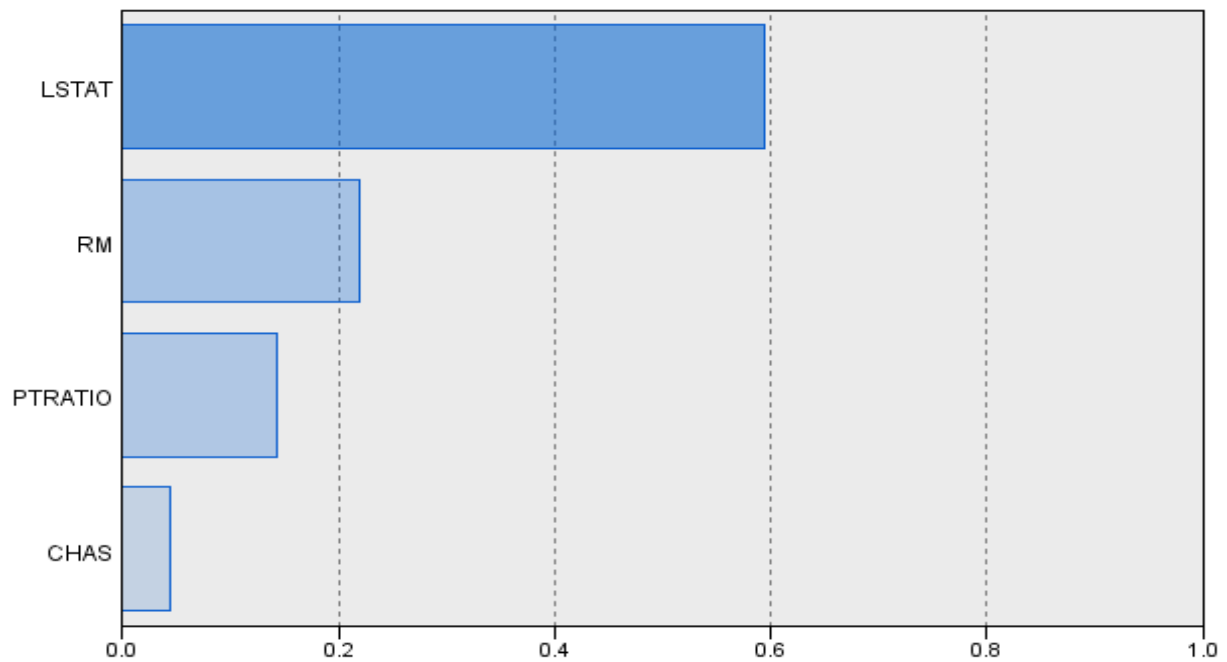
Q2. Car sales

Target Variable = sales

Descriptive Analysis –

	N	Minimum	Maximum	Mean	Std. Deviation
sales	145	.110	230.902	44.51377	46.775129
resale	109	5.860	67.550	18.20381	11.243582
type	145	0	1	.24	.429
price	143	9.699	85.500	27.54871	14.202516
engine_s	144	1.5	5.7	3.025	.9542
horsepow	144	92	345	186.32	52.547
wheelbas	144	92.6	138.7	107.077	6.7486
width	144	65.7	79.9	70.958	3.1959
length	144	152.0	224.2	186.851	12.5928
curb_wgt	143	2.240	5.572	3.36830	.620915
fuel_cap	144	11.9	30.0	17.725	3.5206
mpg	142	15	33	23.91	3.741
Valid N (listwise)	105				

Descriptive Statistics

An initial regression gives us the following details

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	-189.012	143.900		-1.313	.192
	resale	1.725	1.553	.394	1.110	.270
	type	21.845	17.503	.182	1.248	.215
	price	-1.569	1.572	-.438	-.998	.321
	engine_s	36.189	11.616	.675	3.115	.002
	horsepow	-.501	.243	-.534	-2.062	.042
	wheelbas	3.086	1.414	.427	2.182	.032
	width	.012	2.237	.001	.005	.996
	length	.490	.826	.126	.593	.554
	curb_wgt	-67.260	22.491	-.778	-2.991	.004
	fuel_cap	2.434	3.423	.158	.711	.479
	mpg	-.750	2.700	-.055	-.278	.782

The R^2 and R_a^2 value are as follows – 0.376 & 0.301 resp.

Histogram

Dependent Variable: sales

Mean = -2.74E-15
Std. Dev. = 0.946
N = 105
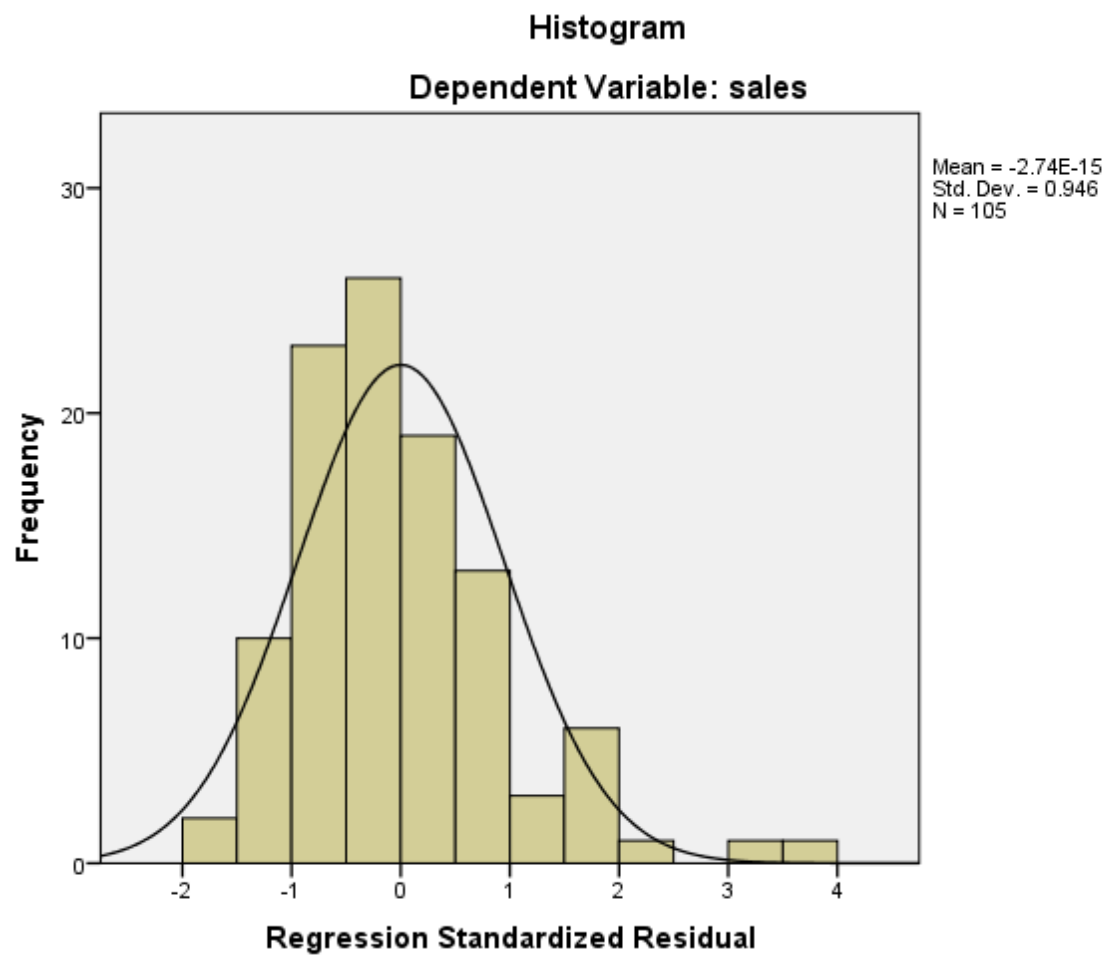
The variables that have a significant impact on the sales of cars are –

- Mpg
  Everyone wants a car with a high mileage and the data agrees with the general census.

Partial Regression Plot
### Dependent Variable: sales
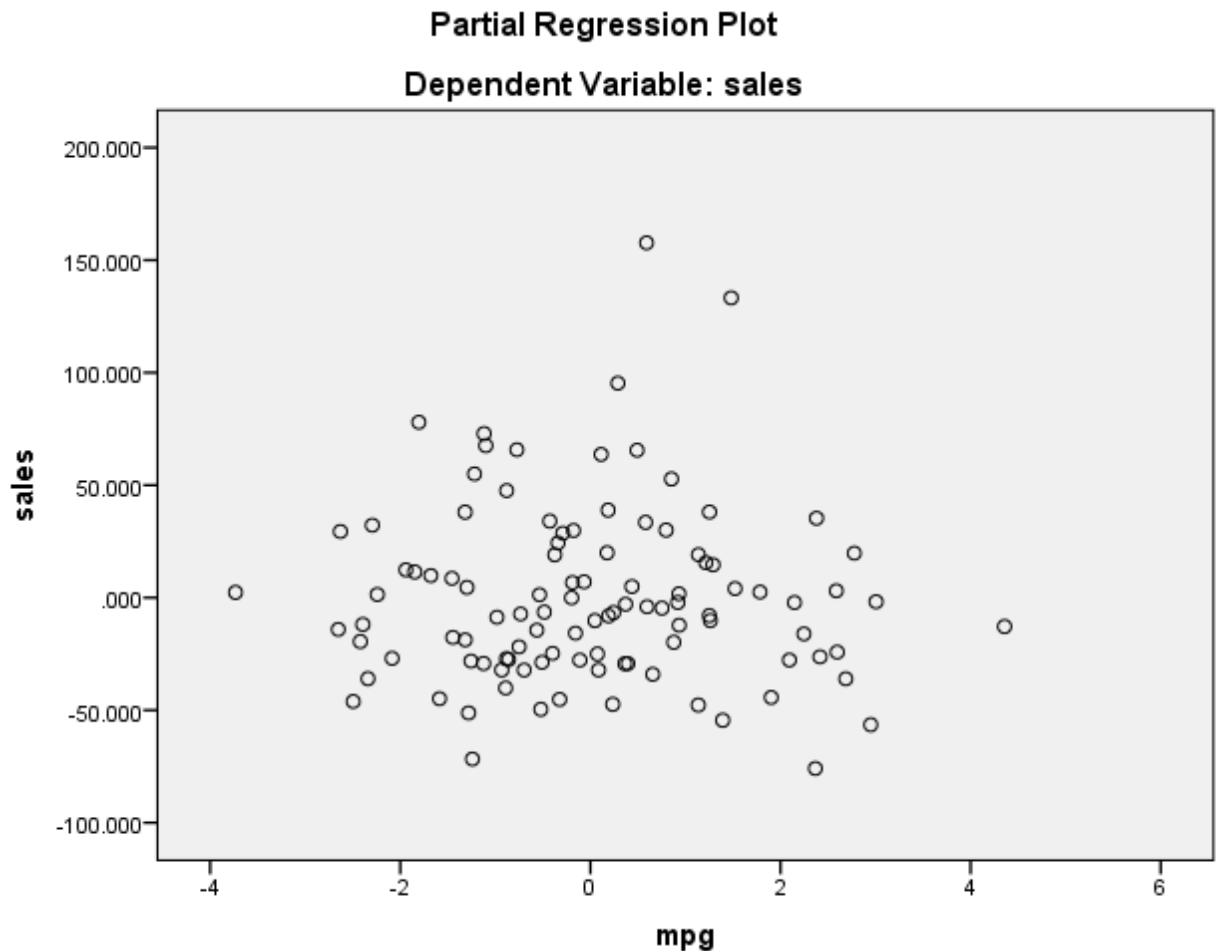
- Fuel Capacity
  The amount of fuel the car can hold is also a significant variable according to the analysis

- Length and Width
  The dimensions of the car play an important role in sales of a car. For example, a family of six may not want to buy a Tata Nano, or someone who lives in a metropolitan city might want a hatchback over a sedan due to parking problems and a higher maneuverability.

- Price
  The pocket friendliness of the car impacts the sales of cars.

Partial Regression Plot

Dependent Variable: sales
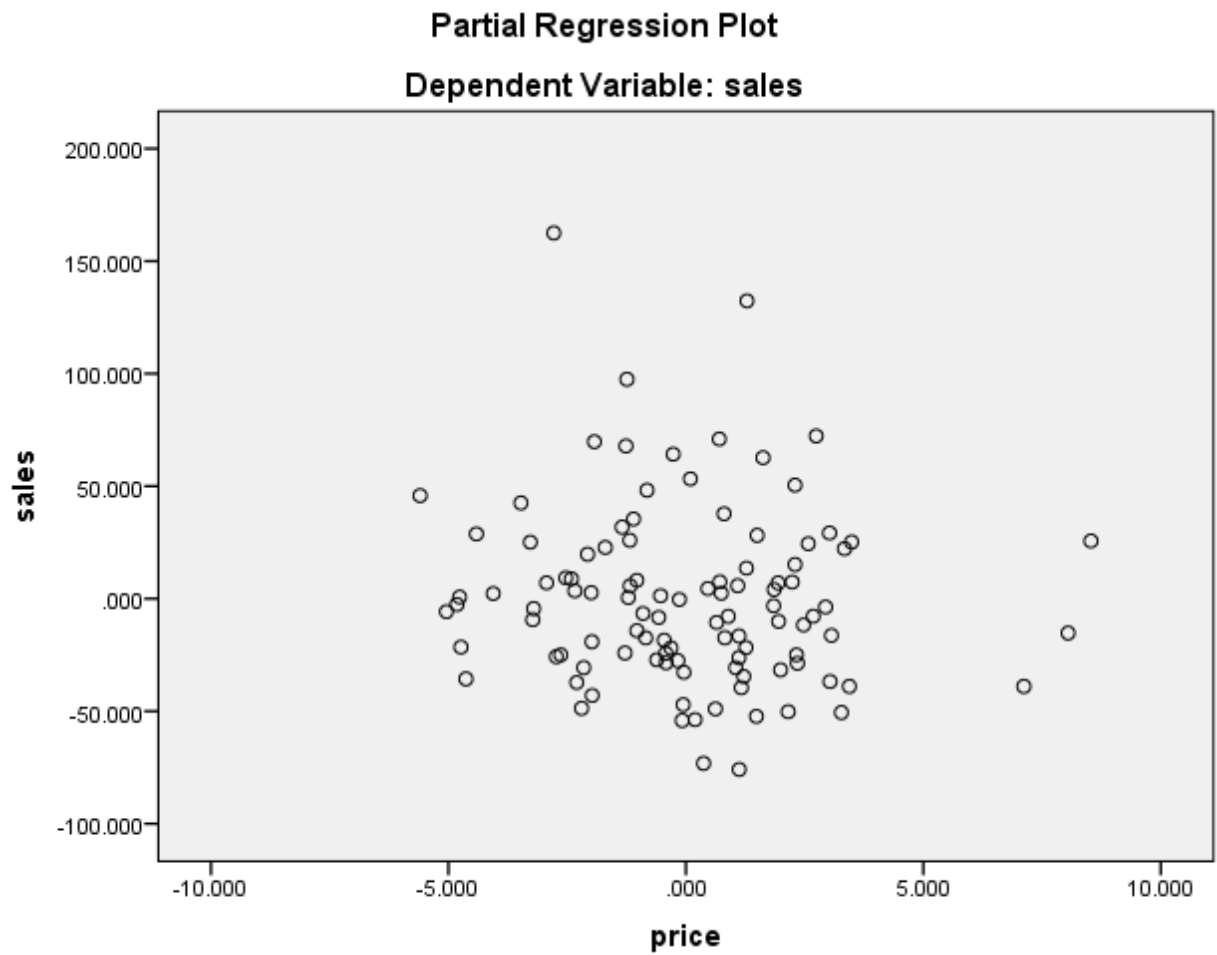
- Resale value
  The resale value of the car also plays a slightly less significant role in the sales of car.

Partial Regression Plot

Dependent Variable: sales
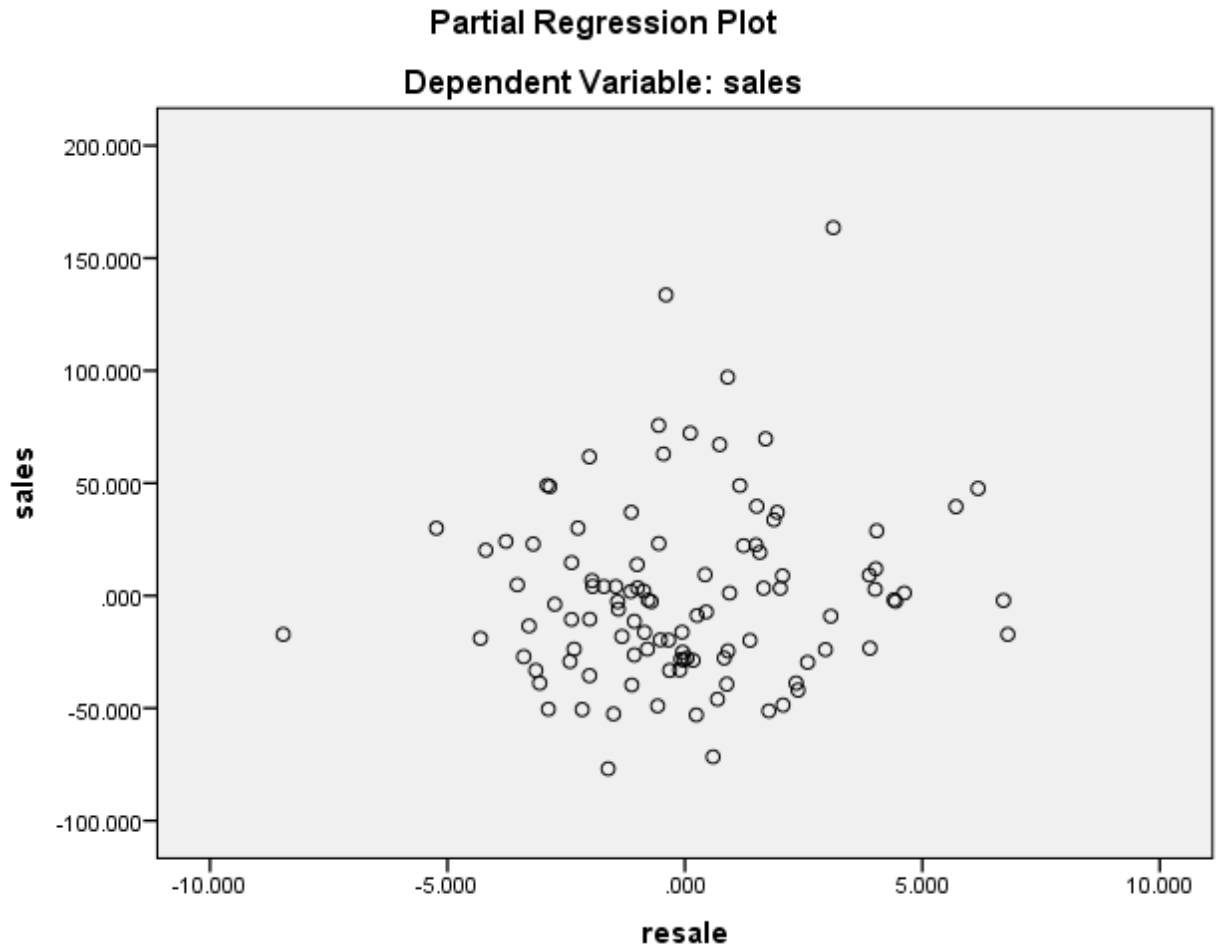
- Manufacturer
  The company manufacturer also plays a role in the sale of cars, either because of a high reputation of the company or customer loyalty.

  Common steps before performing regression for both the questions are
  1. Removal of outliers
  2. Removal of leverage points
  3. Removal of insignificant variables by using AIC values.