TASK1

observation

* According to data, TAX holds the highest values compare than others.
* Similarly, NOX holds the lowest values.

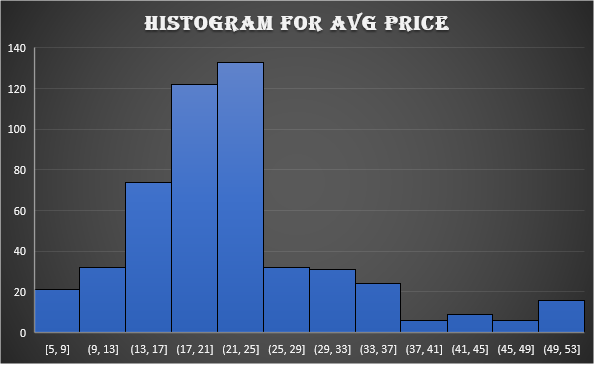
 

TASK2

Observations

* According the data, average price has positive skewness



TASK3

Observations



TASK4



Top 3 positively correlated pairs

1. Tax vs Distance = 0.910228
2. Nox vs Indus = 0.763651447
3. Nox vs Age = 0.73147

Top 3 negatively correlated pairs

1. LSTAT vs average price = -0.73766

2. LSTAT vs Average room = -0.613808272

3. PTRATIO vs Average room = -0.507786686

TASK5



a) What do you infer from the Regression Summary output in terms of variance explained, coefficient value, Intercept, and the Residual plot?

|  |  |
| --- | --- |
| **Column1** | **Column2** |
| R Square | 0.544146298 |
| Intercept | 34.55384088 |
| LSTAT | -0.95004935 |

* R Square value here is about 0.5, it should be near to 1, So it’s not significant
* Coefficient of LSTAT is -0.95005. It is inferred that for each $1000 increase in Average price, there will be a 0.95% decrease in population.
* It is inferred that the Intercept value is 34.5538.
* Residual Plot inferred that all the values are equally distributed.

b)Is LSTAT variable significant for the analysis based on your model?

The p-value for LSTAT variable is 5.08110339438E-88. It is less than 0.05. So, it is inferred that LSTAT variable is significant for the analysis.

TASK6

6) Build a new Regression model including LSTAT and AVG\_ROOM together as independent variables and AVG\_PRICE as dependent variable.

a) Write the Regression equation. If a new house in this locality has 7 rooms (on an average) and has a value of 20 for L-STAT, then what will be the value of AVG\_PRICE? How does it compare to the company quoting a value of 30000 USD for this locality? Is the company Overcharging/ Undercharging?

AVG\_PRICE = Intercept+ (Coefficient of AVG\_ROOM \* value of AVG\_ROOM) + (Coefficient of

LSTAT \* value of LSTAT)

AVG\_PRICE = -1.35827281187456 + (5.09478798433655 \* 7) + (-0.642358334244129 \* 20) AVG\_PRICE = 21.4581

It is inferred that the Average price is $21.4581. But the company quoting a value of 30000 USD for this locality. By the result, it is concluded that the company is overcharging.

R Square = 0.637124475470123 (Qn. 6)

R Square = 0.543241825954707 (Qn. 5)

b) Is the performance of this model better than the previous model you built in Question 5?

Compare in terms of adjusted R-square and explain.

It is inferred that the value of R Square is close to 1, if the count of independent variable

increases.

Based on the analysis, the performance of this model is better than the previous model. (Qn. 5)

TASK7

7) Build another Regression model with all variables where AVG\_PRICE alone be the Dependent

Variable and all the other variables are independent. Interpret the output in terms of adjusted R square, coefficient and Intercept values. Explain the significance of each independent variable with respect to AVG\_PRICE.



For every $1000 of avg. price of houses,

* per capita crime rate by town increases by 0.0487.
* proportion of houses built prior to 1940 increases by 0.03%.
* proportion of non-retail business acres per town increases by 0.13%.
* nitric oxides concentration decreases by 10 million.
* distance from highway increases by 0.2610 miles.
* full-value property-tax rate decreases by 0.0144.
* pupil-teacher ratio by town decreases by 1.0743.
* average number of rooms per house increases by 4.12540.
* lower status (LSTAT) of the population decreases by 0.603%.

TASK 8

8) Pick out only the significant variables from the previous question. Make another instance of the Regression model using only the significant variables you just picked and answer the questions below:

a) Interpret the output of this model.



Adjusted R Square = 0.68868

b) Compare the adjusted R-square value of this model with the model in the previous question,

which model performs better according to the value of adjusted R-square?

Adjusted R Square = 0.6886836818 (Qn.8)

Adjusted R Square = 0.6882986468 (Qn.7)

Adjusted R square for this model is greater comparing to the previous model. So, it is concluded that this model performs better than previous model.

c) Sort the values of the Coefficients in ascending order. What will happen to the average price if the value of NOX is more in a locality in this town?



It is inferred that if the value of NOX is more in a locality in this town, the value of the average price will be reduced.

d) Write the regression equation from this model.

AVG\_PRICE = Intercept + (coefficient of Age \* value of Age) + (coefficient of Indus \* value of Indus) +(coefficient of NOX \* value of NOX) + (coefficient of Distance \* value of Distance) + (coefficient of Tax \* value of Tax) + (coefficient of PTRATIO \* value of PTRATIO) + (Coefficient of Avg\_room \* value of Avg\_room) + (coefficient of LSTAT \* value of LSTAT)