GREAT LEARNING - DA

BUSINESS REPORT - TERRO'S REAL ESTATE AGENCY

GAUTHAM AJAY KANNAN - Batch Jan 23 Online

1. SUMMARY STATISTICS FOR THE GIVEN DATASET

1. **VARIABLE - CRIME_RATE** (per capita crime rate by town) .

MEASURES OF CENTRAL TENDANCY

The mean obtained for the variable crime_rate is around 4.87 which is the average per capita crime rate by town.The median which points to the middle observation is around 4.82 i.e. below 50% of houses or observations have crime rate below 4.82 and remaining 50% have crime rate above 4.82. The mode which refers to the most frequently occurred value in the observations is 3.43.

MEASURES OF DISPERSION

The range which typically refers to the difference between the maximum and minimum value is 9.95. The standard deviation which is the squared average deviation is around 2.92.

KURTOSIS AND SKEWNESS

Kurtosis which refers to the sharpness of the peak of normal distribution is around -1.189 which is negative in nature .Hence the peak will be flat.Skewness is a measure of asymmetry of data .Here ,it is around 0.02 which is slightly positive in nature.

2. **VARIABLE - AGE** (proportion of houses built prior to 1940 (in percentage terms)) .

MEASURES OF CENTRAL TENDANCY

The mean obtained for the variable **age** is around 68.57% which denotes that 68.57% of houses were built prior to 1940.The median which points to the middle observation is around 77.5% i.e. below 50% of houses or observations have age percentage below 77.5% and remaining have age percentage above 77.5. The mode which refers to the most frequently occurred value in the observations is 100%.

MEASURES OF DISPERSION

The range which typically refers to the difference between the maximum and minimum value is 100%. The standard deviation which is the squared average deviation is around 28.15%.

KURTOSIS AND SKEWNESS

Kurtosis which refers to the sharpness of the peak of normal distribution is around -0.968 which is negative in nature .Hence the peak will be flat.Skewness is a measure of asymmetry of data .Here ,it is around -0.60 which is negative in nature.Hence the observations are trailing off towards left.

3. **VARIABLE - INDUS** (proportion of non-retail business acres per town (in percentage terms)) .

MEASURES OF CENTRAL TENDANCY

The mean obtained for the variable **indus** is around 11.13% which denotes that each houses are having an average value of 11.13% in their respective localities .The median which points to the middle observation is around 9.69 % i.e. below 50% of houses or observations have indus value below 9.69% and remaining have indus percentage above 9.69%. The mode which refers to the most frequently occurred value in the observations is 18.1%.

MEASURES OF DISPERSION

The range which typically refers to the difference between the maximum and minimum value is 27.28%. The standard deviation which is the squared average deviation is around 6.86%.

KURTOSIS AND SKEWNESS

Kurtosis which refers to the sharpness of the peak of normal distribution is around -1.233 which is negative in nature .Hence the peak will be flat.Skewness is a measure of asymmetry of data .Here ,it is around 0.29 which is positive in nature.Hence the observations or houses are trailing off towards right side of the central value.

4. **VARIABLE - NOX** (nitric oxides concentration (parts per 10 million)) .

MEASURES OF CENTRAL TENDANCY

The mean obtained for the variable **NOX** is around 0.55 which denotes that the houses in the given dateset have an average NOX value 0.55 .The median which points to the middle observation is around 0.538 i.e. below 50% of houses or observations have NOX concentration below 0.538 and remaining have NOX concentration above 0.538. The mode which refers to the most frequently occurred value in the observations is 0.538.Hence, majority of the houses or observations have NOX concentration around 0.538 within their locality.

MEASURES OF DISPERSION

The range which typically refers to the difference between the maximum and minimum value is 0.486. The standard deviation which is the squared average deviation is around 0.1158.

KURTOSIS AND SKEWNESS

Kurtosis which refers to the sharpness of the peak of normal distribution is around -0.064 which is negative in nature .Hence the peak will be flat.Skewness is a measure of asymmetry of data .Here ,it is around 0.729 which is positive in nature.Hence the observations are trailing off towards right of the central or average value.

5. **VARIABLE - distance** (distance from highway (in miles)) .

MEASURES OF CENTRAL TENDANCY

The mean obtained for the variable **distance** is around 9.54 which denotes that the houses in the given dateset have an average distance of 9.54 from the highway .The median which points to the middle observation is around 5 i.e. below 50% of houses or observations have distance below 5 miles and remaining have distance above 5 miles. The mode which refers to the most frequently occurred value in the observations is 24 miles.Hence, majority of the houses or observations have a distance around 24 miles from the highway.

MEASURES OF DISPERSION

The range which typically refers to the difference between the maximum and minimum value is 23 miles. The standard deviation which is the squared average deviation is around 8.7 miles.

KURTOSIS AND SKEWNESS

Kurtosis which refers to the sharpness of the peak of normal distribution is around -0.867which is negative in nature .Hence the peak will be flat.Skewness is a measure of asymmetry of data .Here ,it is around 1.004 which is positive in nature.Hence the observations are trailing off towards right of the central or average value.

6. **VARIABLE - avg_room** (average number of rooms per house) .

MEASURES OF CENTRAL TENDANCY

The mean obtained for the variable **avg_room** is around 6 rooms which denotes that each house in the given dateset have an average of 6 rooms .The median which points to the middle observation is around 6 i.e. below 50% of houses or observations have 6 rooms and remaining have distance above 6 miles. The mode which refers to the most frequently occurred value in the observations is 6 rooms.Hence, majority of houses or observations 6 rooms.

MEASURES OF DISPERSION

The range which typically refers to the difference between the maximum and minimum value is 6 rooms. The standard deviation which is the squared average deviation is around 1 room.

KURTOSIS AND SKEWNESS

Kurtosis which refers to the sharpness of the peak of normal distribution is around 1.89 which is positive in nature .Hence the peak will be sharp in nature.Skewness is a measure of asymmetry of data .Here ,it is around 1.89 which is positive in nature.Hence the observations are trailing off towards right of the central or average value.

7. **VARIABLE - LSTAT** (% lower status of the population) .

MEASURES OF CENTRAL TENDANCY

The mean obtained for the variable **LSTAT** is around 12.6% which denotes that each house in the given dateset have an average LSTAT value or percentage that contributes to the lower status of population is around 12.6% .The median which points to the middle observation is around 11.36% i.e. below 50% of houses or observations have LSTAT value above 11.36% and remaining have LSTAT value below 11.36%. The mode which refers to the most frequently occurred value in the observations is 5.7% .Hence, majority of houses have 5.7% of the lower status of the population in the town.

MEASURES OF DISPERSION

The range which typically refers to the difference between the maximum and minimum value is 5.2%. The standard deviation which is the squared average deviation is around 0.7%.

KURTOSIS AND SKEWNESS

Kurtosis which refers to the sharpness of the peak of normal distribution is around 1.89 which is positive in nature .Hence the peak will be sharp in nature.Skewness is a measure of asymmetry of data .Here ,it is around 0.403 which is nearly positive in nature.Hence the observations are slightly trailing off towards right of the central or average value.

8. **VARIABLE - avg_price** (average value of houses in $1000) .

MEASURES OF CENTRAL TENDANCY

The mean obtained for the variable **avg_price** is around $22530 which denotes that each house in the given dateset have an average price value of $ 22530 .The median which points to the middle observation is around $21200 i.e. below 50% of houses or observations have price above $21200 and remaining have price below $21200. The mode which refers to the most frequently occurred value in the observations is $50000.Hence, majority of houses or observations price of $50000.

MEASURES OF DISPERSION

The range which typically refers to the difference between the maximum and minimum value is $45000 . The standard deviation which is the squared average deviation is around $9000.

KURTOSIS AND SKEWNESS

Kurtosis which refers to the sharpness of the peak of normal distribution is around 1.495 which is positive in nature .Hence the peak will be sharp in nature.Skewness is a measure of asymmetry of data .Here ,it is around 1.108 which is positive in nature.Hence the observations are trailing off towards right of the central or average value.

9. **VARIABLE - tax** (full value property tax rate per $10000) .

MEASURES OF CENTRAL TENDANCY

The mean obtained for the variable tax is around 408.2 which denotes that each house in the given dateset have an average tax value of 408.2 .The median which points to the middle observation is around 330 i.e. below 50% of houses or observations have tax value below 330 and remaining have tax value above 330. The mode which refers to the most frequently occurred value in the observations is 666 rooms.Hence, majority of houses or observations have tax value of 666.

MEASURES OF DISPERSION

The range which typically refers to the difference between the maximum and minimum value is 524 . The standard deviation which is the squared average deviation is around 168.5.

KURTOSIS AND SKEWNESS

Kurtosis which refers to the sharpness of the peak of normal distribution is around -1.142 which is negative in nature .Hence the peak will be flat in nature.Skewness is a measure of asymmetry of data .Here ,it is around 0.6699 which is positive in nature.Hence the observations are trailing off towards right of the central or average value.

10. **VARIABLE - PTRARIO** (pupil teacher ratio by town) .

MEASURES OF CENTRAL TENDANCY

The mean obtained for the variable **PTRATIO** is around 18.45 which denotes that each house in the given dateset have an average PTRATIO of 18.45 .The median which points to the middle observation is around 19.05 i.e. below 50% of houses or observations have PTRATIO below 19.05 and remaining have PTRATIO above 19.05. The mode which refers to the most frequently occurred value in the observations is 20.2.Hence, majority of houses or observations have PTRATIO of 20.2 in their respective localities.

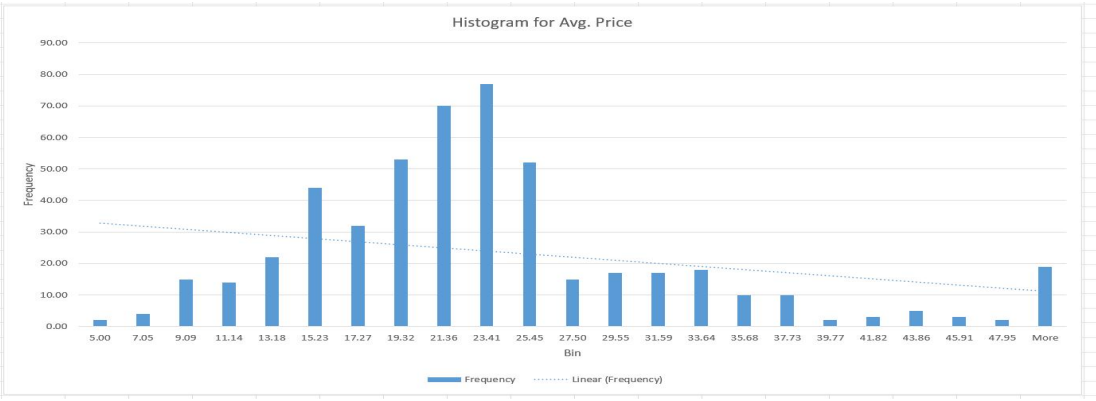
MEASURES OF DISPERSION

The range which typically refers to the difference between the maximum and minimum value is 9.4 . The standard deviation which is the squared average deviation is around 2.164.

KURTOSIS AND SKEWNESS

Kurtosis which refers to the sharpness of the peak of normal distribution is around -0.285 which is slightly negative in nature .Hence the peak will be flat in nature.Skewness is a measure of asymmetry of data .Here ,it is around -0.802 which is negative in nature.Hence the observations are trailing off towards left side of the central or average value.

2. HISTOGRAM OF THE VARIABLE - AVG_PRICE



The above figure shows us the distribution of the observations or houses in Boston based on their average price .From the figure ,we could see that there is an gradual increase from left to right and most of the houses in Boston have average prices distributed in the middle of the diagram and it is within the range $19320 - $ $25450 with $234100 being the highest denoting that majority of houses have an average price of $234100.Then there is an major difference in the count between the bins representing the values $25450 and $27500.

3. COVARIANCE MATRIX

COVARIANCE MATRIX - TERRO'S REAL ESTATE AGENCY

	CRIME_RATE	AGE	INDUS	NOX	DISTANCE	TAX	PTRATIO	AVG_ROOM	LSTAT	AVG_PRICE
CRIME_RATE	8.516147873									
AGE	0.562915215	790.7924728								
INDUS	-0.110215175	124.2678282	46.97142974							
NOX	0.000625308	2.381211931	0.605873943	0.013401099						
DISTANCE	-0.229860488	111.5499555	35.47971449	0.615710224	75.66653127					
TAX	-8.229322439	2397.941723	831.7133331	13.02050236	1333.116741	28348.6236				
PTRATIO	0.068168906	15.90542545	5.680854782	0.047303654	8.74340249	167.8208221	4.677726296			
AVG_ROOM	0.056117778	-4.74253803	-1.884225427	-0.024554826	-1.281277391	-34.51510104	-0.539694518	0.492695216		
LSTAT	-0.882680362	120.8384405	29.52181125	0.487979871	30.32539213	653.4206174	5.771300243	-3.073654967	50.89397935	
AVG_PRICE	1.16201224	-97.39615288	-30.46050499	-0.454512407	-30.50083035	-724.8204284	-10.09067561	4.484565552	-48.35179219	84.41955616

Covariance refers to the joint dispersion of two variables.Two variables X and Y are said to have positive covariance if their values both are mostly above or below their averages.Two variables X and Y are said to have negative covariance if their values both are mostly on opposite side of their averages.From the above observations, we could see for the variable **crime_rate** has negative covariance with most of the other variables like INDUS,NOX,TAX and LSTAT when compared with other variables.

4. CORRELATION MATRIX

CORRELATION MATRIX - TERRO'S REAL ESTATE AGENCY

	CRIME_RATE	AGE	INDUS	NOX	DISTANCE	TAX	PTRATIO	AVG_ROOM	LSTAT	AVG_PRICE
CRIME_RATE	1									
AGE	0.006859463	1								
INDUS	-0.005510651	0.644778511	1							
NOX	0.001850982	0.731470104	0.763651447	1						
DISTANCE	-0.009055049	0.456022452	0.595129275	0.611440563	1					
TAX	-0.016748522	0.506455594	0.72076018	0.6680232	0.910228189	1				
PTRATIO	0.010800586	0.261515012	0.383247556	0.188932677	0.464741179	0.460853035	1			
AVG_ROOM	0.02739616	-0.240264931	-0.391675853	-0.302188188	-0.209846668	-0.292047833	-0.355501495	1		
LSTAT	-0.042398321	0.602338529	0.603799716	0.590878921	0.488676335	0.543993412	0.374044317	-0.613808272	1	
AVG_PRICE	0.043337871	-0.376954565	-0.48372516	-0.427320772	-0.381626231	-0.468535934	-0.507786686	0.695359947	-0.73766273	1
POSITIVE CORRELATION	0.043337871	0.731470104	0.763651447	0.6680232	0.910228189	0.543993412	0.374044317	0.695359947		
NEGATIVE CORRELATION	-0.042398321	-0.376954565	-0.48372516	-0.427320772	-0.381626231	-0.468535934	-0.507786686	-0.613808272	-0.73766273	

TOP 3 POSITIVELY CORRELATION PAIRS	DISTANCE VS TAX	0.910228189	TOP 3 NEGATIVE	LSTAT VS AVG_PRICE	-0.737662726
	INDUSTRY VS NOX	0.763651447		AVG_ROOM VS LSTAT	-0.613808272
	AGE VS NOX	0.731470104		PTRATIO VS AVG_PRICE	-0.507786686

Correlation or correlation coefficient is an dimensionless version of covariance.It helps us identifying the relationship between two variables.Relationship could be classified into positive,negative and zero .From the above figure we could see the top three positive and negative correlated pairs.

POSITIVE CORRELATED PAIRS

There exists a positive linear relationship between two variables i.e. if one variable is found to be increasing gradually,the second variable is also increasing correspondingly.The top three positive correlated pairs are

1. DISTANCE (distance from highway) VS TAX (full value property tax rate per $10000
2. INDUSTRY (proportion of non retail business acres per town) VS NOX (nitric oxide concentration).
3. AGE (proportion of houses built prior to 1940) VS NOX (nitric oxide concentration).

NEGATIVE CORRELATED PAIRS

There exists a negative linear relationship between two variables i.e. if one variable is found to be increasing gradually,the second variable is decreasing correspondingly or vice versa.The top three negative correlated pairs are

1. LSTAT (% lower status of population) VS AVG_PRICE (avg. value of houses)
2. AVG_ROOM (average number of rooms per house) VS LSTAT (%lower status of

population).

3.PTRATIO (pupil teacher ratio by town) vs AVG_PRICE(avg. value of houses).

5.SINGLE LINEAR REGRESSION MODEL USING LSTAT AS A INDEPENDENT VARIABLE

SUMMARY OUTPUT

SUMMARY OUTPUT SINGLE LINEAR REGRESSION MODEL USING *LSTAT* AS INDEPENDANT VARIABLE AND *AVG_PRICE* AS DEPENDANT VARIABLE

Regression Statistics	
Multiple R	0.737662726
R Square	0.544146298
Adjusted R Square	0.543241826
Standard Error	6.215760405
Observations	506

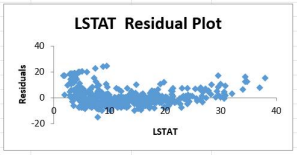
ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	1	23243.914	23243.914	601.6178711	5.0811E-88
Residual	504	19472.38142	38.63567742		
Total	505	42716.29542			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	34.55384088	0.562627355	61.41514552	3.7431E-236	33.44845704	35.65922472	33.44845704	35.65922472
LSTAT	-0.950049354	0.038733416	-24.52789985	5.0811E-88	-1.0261482	-0.873950508	-1.0261482	-0.873950508

RESIDUAL OUTPUT

Observation	*Predicted AVG_PRICE*	*Residuals*
1	29.8225951	-5.822595098
2	25.87038979	-4.270389786
3	30.72514198	3.974858016
4	31.76069578	1.639304221
5	29.49007782	6.709922176
6	29.60408375	-0.904083746
7	22.74472741	0.155272588



Single linear regression helps us in finding a linear relationship or a line that fits or represents all the observations between the dependant variable (y) and the independent variable x.Here, we have created a single linear regression model with AVG_PRICE as dependant variable and LSTAT as independent variable.The R square value which represents the variance of the model is found to be around 0.54 which has 0.5 difference between one. The following regression equation is formed using the given values.

Intercept value is 34.55 and the minimum value or the starting point which the dependant variable could obtain when there is 0% LSTAT is 34.55 .The minimum price would be $34550.The coefficient value of the LSTAT variable is -0.95 which is negative in nature.When it comes to the significance ,we have to consider the p^2 value which in this case ,it is 3.7431E-236 for LSTAT variable .

$$Y= 34.55 -0.95X$$

A variable is said to be significant when its p^2 value is less than 0.05 .Hence, the variable LSTAT is a significant predictor of the dependant variable i.e. the average price of the house. From the residual plot of the LSTAT variable ,we couldn't see any pattern or any relationship between its value and their corresponding residual value.

6. MULTI LINEAR REGRESSION MODEL USING LSTAT AND AVG_ROOM AS A INDEPENDENT VARIABLES

SUMMARY OUTPUT

SUMMARY OUTPUT FOR MULTI LINEAR REGRESSION MODEL USING *AVG_ROOM* AND *LSTAT* AS DEPENDANT VARUABLES AND *AVG_PRICE* AS DEPENDANT VARIABLE

Regression Statistics	
Multiple R	0.799100498
R Square	0.638561606
Adjusted R Square	0.637124475
Standard Error	5.540257367
Observations	506

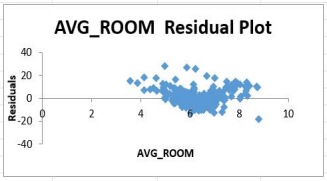
ANOVA

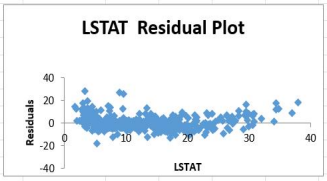
	df	*SS*	*MS*	*F*	*Significance F*
Regression	2	27276.98621	13638.49	444.3308922	7.0085E-112
Residual	503	15439.3092	30.69445		
Total	505	42716.29542			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	-1.358272812	3.17282778	-0.4281	0.668764941	-7.591900282	4.875354658	-7.591900282	4.875354658
AVG_ROOM	5.094787984	0.4444655	11.46273	3.47226E-27	4.221550436	5.968025533	4.221550436	5.968025533
LSTAT	-0.642358334	0.043731465	-14.6887	6.66937E-41	-0.728277167	-0.556439501	-0.728277167	-0.556439501

RESIDUAL OUTPUT

Observation	*Predicted AVG_PRICE*	*Residuals*
1	28.94101368	-4.941013681
2	25.48420566	-3.884205661
3	32.65907477	2.040925231
4	32.40652	0.99348
5	31.63040699	4.569593009
6	28.05452701	0.645472994
7	21.28707846	1.612921545





Multi Linear Regression is formed when we are trying to establish an relationship or pattern between one dependant variable and two or more independent variables or predictors.Here ,we have created an multi linear regression model using AVG_PRICE as dependant variable and LSTAT and AVG_ROOM as independent variables.The regression equation formed by the created model is

$$Y = -1.358+5.09X_0 -0.642X_1$$

According to the given question ,if a new house has 7 rooms and has an LSTAT value of 20,then the predicted value based on the above regression equation is $21500 .The company quotes a value of $30000.The difference between the predicted and the provided value is $8500 .Hence,the company is overcharging.

By comparing the adjusted R-square values of this and the previous model,we could see the R square value of the current model (0.637) is greater than the previous one (0.543).Hence,by this comparison ,we could say that the current model is better than the previous one.

7. MULTI LINEAR REGRESSION MODEL USING ALL THE INDEPENDENT VARIABLES

SUMMARY OUTPUT FOR MULTI LINEAR REGRESSION MODEL USING ALL DEPENDANT VARIABLES - TERRO'S REAL ESTATE AGENCY

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.832978824
R Square	0.69385372
Adjusted R Square	0.688298647
Standard Error	5.1347635
Observations	506

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	9	29638.8605	3293.206722	124.9045049	1.9328E-121
Residual	496	13077.43492	26.3657962		
Total	505	42716.29542			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	29.24131526	4.817125596	6.070282926	2.53978E-09	19.77682784	38.70580267	19.77682784	38.70580267
CRIME_RATE	0.048725141	0.078418647	0.621346369	0.534657201	-0.105348544	0.202798827	-0.105348544	0.202798827
AGE	0.032770689	0.013097814	2.501996817	0.012670437	0.00703665	0.058504728	0.00703665	0.058504728
INDUS	0.130551399	0.063117334	2.068392165	0.03912086	0.006541094	0.254561704	0.006541094	0.254561704
NOX	-10.3211828	3.894036256	-2.650510195	0.008293859	-17.97202279	-2.670342809	-17.97202279	-2.670342809
DISTANCE	0.261093575	0.067947067	3.842602576	0.000137546	0.127594012	0.394593138	0.127594012	0.394593138
TAX	-0.01440119	0.003905158	-3.687736063	0.000251247	-0.022073881	-0.0067285	-0.022073881	-0.0067285
PTRATIO	-1.074305348	0.133601722	-8.041104061	6.58642E-15	-1.336800438	-0.811810259	-1.336800438	-0.811810259
AVG_ROOM	4.125409152	0.442758999	9.317504929	3.89287E-19	3.255494742	4.995323561	3.255494742	4.995323561
LSTAT	-0.603486589	0.053081161	-11.36912937	8.91071E-27	-0.70777824	-0.499194938	-0.70777824	-0.499194938

The equation formed by the above regression model using all the independent variables is

The adjusted R square is 0.688 which is closer to one and greater when comparing it with the pre

$$Y=29.24+O.O4X_0+0.03X_1+0.13X_2-10.32X_3+0.26X_4-0.01X_5-1.07X_6+4.12X_7-0.6X_8$$

or the price of the house as per the model is $29240.The significance of each independent variables are as follows.

VARIABLES	*P-value*	*SIGNIFICANT(P-value<0.05)*
CRIME_RATE	0.534657201	FALSE
AGE	0.012670437	TRUE
INDUS	0.03912086	TRUE
NOX	0.008293859	TRUE
DISTANCE	0.000137546	TRUE
TAX	0.000251247	TRUE
PTRATIO	6.58642E-15	TRUE
AVG_ROOM	3.89287E-19	TRUE
LSTAT	8.91071E-27	TRUE

From the above figure ,we could see that except for crime rate (which is greater than 0.05),the remaining independent variables used in the model are found to be significant predictor of the average price of the house.

8. MULTI LINEAR REGRESSION MODEL USING ALL THE SIGNIFICANT VARIABLES

MULTI LINEAR REGRESSION MODEL USING ALL THE SIGNIFICANT VARIABLES - TERRO'S REAL ESTATE AGENCY

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.832835773
R Square	0.693615426
Adjusted R Square	0.688683682
Standard Error	5.131591113
Observations	506

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	8	29628.68	3703.585	140.6430411	1.911E-122
Residual	497	13087.61	26.33323		
Total	505	42716.3			

	Coefficients	*andard Erri*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	29.42847349	4.804729	6.124898	1.84597E-09	19.98838959	38.8685574	19.98838959	38.8685574
AGE	0.03293496	0.013087	2.516606	0.012162875	0.007222187	0.058647734	0.007222187	0.058647734
INDUS	0.130710007	0.063078	2.072202	0.038761669	0.006777942	0.254642071	0.006777942	0.254642071
NOX	-10.27270508	3.890849	-2.64022	0.008545718	-17.9172457	-2.628164466	-17.9172457	-2.628164466
DISTANCE	0.261506423	0.067902	3.851242	0.000132887	0.128096375	0.394916471	0.128096375	0.394916471
TAX	-0.014452345	0.003902	-3.70395	0.000236072	-0.022118553	-0.006786137	-0.022118553	-0.006786137
PTRATIO	-1.071702473	0.133454	-8.03053	7.08251E-15	-1.333905109	-0.809499836	-1.333905109	-0.809499836
AVG_ROOM	4.125468959	0.442485	9.3234	3.68969E-19	3.256096304	4.994841615	3.256096304	4.994841615
LSTAT	-0.605159282	0.05298	-11.4224	5.41844E-27	-0.70925186	-0.501066704	-0.70925186	-0.501066704

The above model is created using only the significant variables that are identified from the previous model .The regression equation formed by this model is

$$Y=29.22+0.03X_0+0.13X_1-10.27X_2+0.26X_3-0.01X_4-1.07X_5+4.13X_6-0.6X_7$$

The value of the intercept is $29220 i.e. the minimum price of the house that could be predicted from the model when the dependant variables are found to be minimum or negligible is $29220.When comparing the R square value of the current model (0.689) with the previous one (0.688),the current one is slightly better than the previous model but there is no greater difference in the value. The ascending order of the coefficients are as follows.

Coefficients	
NOX	-10.27270508
PTRATIO	-1.071702473
LSTAT	-0.605159282
TAX	-0.014452345
AGE	0.03293496
INDUS	0.130710007
DISTANCE	0.261506423
AVG_ROOM	4.125468959

From the correlation matrix ,we could see that there exists an negative linear relationship between the price of the house and the NOX concentration i.e. if the NOX concentration increases ,the price decreases or vice versa.

ATTACHMENTS

Attaching the excel sheet for the reference.

TERRO'S_REA_EX
CEL_GAUTHAM_A.