Practical No. 10 : Multiple Linear Regression

The assessed value (in lakh rupees), the size of the house (thousands square feet) and the age of houses (in years) is given in following table.

House No.	Market Price (Rs 1,00,000)	Size of the house (Thousands of square feet)	Age of house (years)
1	63.0	1.605	35
2	65.1	2.489	45
3	69.9	1.553	20
4	76.8	2.404	32
5	73.9	1.884	25
6	77.9	1.558	14
7	74.9	1.748	8
8	78.0	3.105	10
9	79.0	1.682	28
10	83.4	2.470	30
11	79.5	1.820	2
12	83.9	2.143	6
13	79.7	2.121	14
14	84.5	2.485	9
15	96.0	2.300	19
16	109.5	2.714	4
17	102.5	2.463	5
18	121.0	3.076	7
19	104.9	3.048	3
20	128.0	3.267	6
21	129.0	3.069	10
22	117.9	4.765	11
23	140.0	4.540	8

Enter this data in Minitab and generate the following reports:

Questions :

a) Fit the multiple regression equation of market price of house on size of the house and age of the house
b) Interpret the meaning of the slopes
c) Compute the coefficient of multiple determination and interpret the meaning of value
d) Determine the adjusted R^2 interpret the meaning of value
e) Determine whether there is a significant relationship between market price of house and the two independent variables (size and age) at the 0.05 level of significance.
f) At the 0.05 level of significance, determine whether each independent variable makes a significant contribution to the regression model.

g) Perform residual analysis for checking model assumptions.

Solution :

Step 1 : Type your data into the data pane of a worksheet. Make sure you put your data into columns. Use column header for “House No.”, “Market Price”, “Size of House” and “Age of House”. Type the “House No.” data into column C1, “Market Price” data into column C2, “Size of House” data into column C3 and “Age of House” data into column C4.

Step 2 : To perform multiple linear regression, under the drop-down menu “STAT, choose “Regression” then “Regression” then “Fit Regression Model…”. A “Regression” dialogue box will appear. Set the “Responses:” as “C2 Market Price” and “Continuous Predictors:” as “C3 Size of House” and “C4 Age of House” from the table on the left. Click “OK”.

Step 3 : Click on the “Storage…” option. A “Regression: Storage” dialogue box will appear. Check the “Fits” and “Residuals” checkboxes. Click “OK”.

Step 4 : Click on the “Graphs…” option. A “Regression: Graphs” dialogue box will appear. Check the “Four in one” radio box. Click “OK”.

Y = Market price of house (per lakh rupees)
X1 = Size of the house (per thousand square feet)
X2 = Age of the house (years)

Regression Equation :

Market Price = 57.4 + 17.72 Size of House - 0.666 Age of House

Coefficients :

Term	Coef	SE Coef	T-Value	P-Value	VIF
Constant	57.4	10.0	5.73	0.000	
Size of House	17.72	3.15	5.63	0.000	1.12
Age of House	-0.666	0.228	-2.92	0.008	1.12

Conclusion:

(i) Sample intercept is 57.4 with a standard error of 10. The sample intercept is significant (p-value = 0.000).
(ii) Sample slope of market price of house on size of the house, controlling house age is 17.72 with standard error of 3.15. The slope is significant (p-value = 0.000).
(iii) Sample slope of market price of house on age of the house, controlling house size is – 0.666 with standard error of 0.228. The slope is significant (p-value = 0.008).
(iv) Both house size and house age are significant and hence retained in the model, but house size is contributing largely on market price of house as compared to house age (on observing slopes).

(v) The meaning of the slope b1 = 17.72 is that if house size increases by 1 unit (i.e. 1 thousand square feet), the house price increases in average by an amount of Rs 17.72 (x 100000) = Rs 17,72,000, controlling house age.

(vi) The meaning of the slope b2 = -0.666 is that if house age increases by 1 unit (i.e. 1 year), the house price decreases in average by an amount of Rs 0.666 (x 100000) = Rs 66,600, controlling house size.

Model Summary :

S	R-sq	R-sq(adj)	R-sq(pred)
11.9604	74.11%	71.52%	65.49%

Conclusion :

(i) The adjusted coefficient of determination shows that 71.52 % variation in the market price of house is due to linear relationship of market price of house, size of the house and age of the house i.e. linear regression of market price of house, size of the house and age of the house. The remaining 28.48 % variation in market price of house is due to factors other than house size and house price. Hence, the reliability of the fitted equation is high i.e. we can reasonably predict house price knowing house size and house age.

(ii) The standard error of estimate is 11.96

F-test for whole model (Reliability of whole model) :

Analysis of Variance:

Source	DF	Adj SS	Adj MS	F-Value	P-Value
Regression	2	8190	4094.9	28.63	0.000
Error	20	2861	143.1		
Total	22	11051			

Conclusion:

The ANOVA table shows that the whole model is significant (p-value = 0.000).

Residual Analysis :

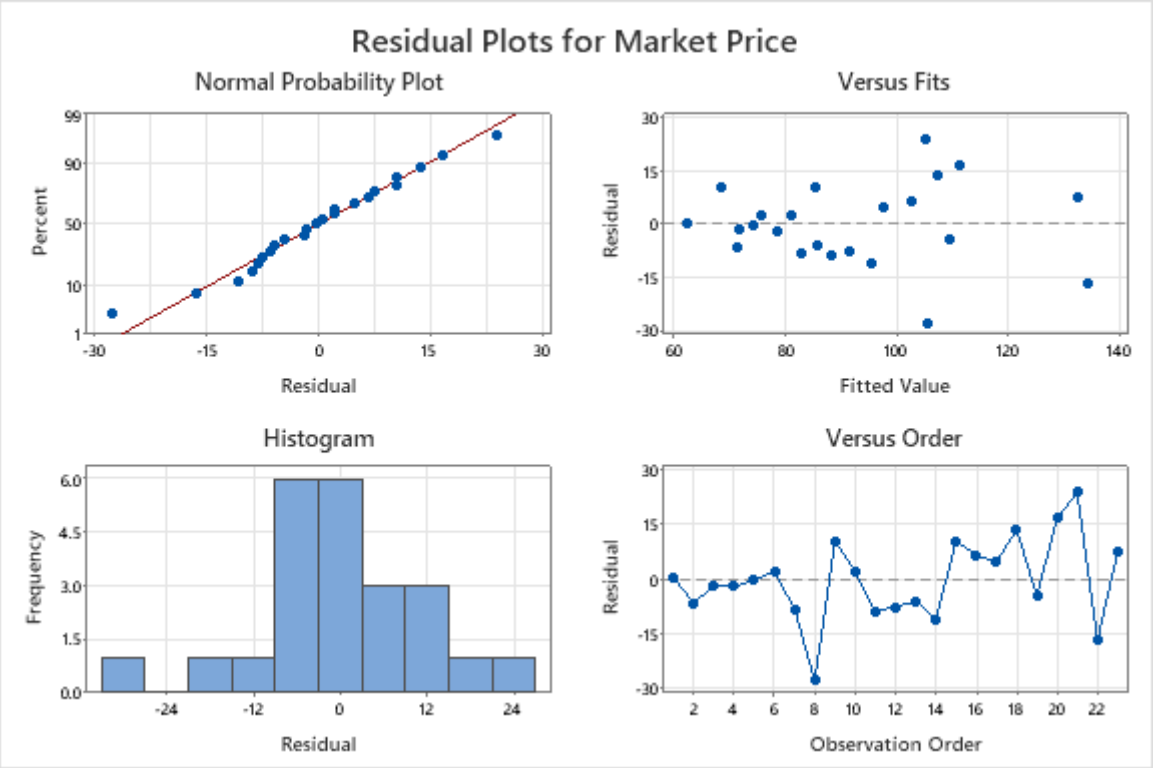

Fig 1 : Residual Analysis

(i) **Normality check of error distribution**: The histogram and normal probability plot shows that the error distribution is not satisfactorily normal i.e. it is right skewed. It may be due to the small sample size.

(ii) **Equal variance check**: The second graph shows that the distribution of dots about reference line e = 0 is almost same or equal, suggesting homoscedasticity. Hence, the assumption of equal variance is satisfactorily met, however there are few extreme values in the graph.

(iii) **Linear relationship check**: The second graph also shows that the distribution of dots about the reference line e = 0 (below and above the line) has no obvious pattern (pattern is random), indicating linear relationship between house price, house size and house age.

(iv) **Independence of error check**: The fourth graph (graph of error vs observation order) shows that error distribution is in random pattern, indicating independence of errors. It means that positive and negative errors are random order.

Fits and Diagnostics for Unusual Observations :

Obs	Market Price	Fit	Resid	Std Resid	
8	78.00	105.70	-27.70	-2.40	R
21	129.00	105.06	23.94	2.07	R

R Large residual

Conclusion :

The observations no. 8 and 21 are unusual observations since their residual are quite high, and they should be considered for removal from the data set, in order to improve the reliability of the fitted equation.

Worksheet :

	C1	C2	C3	C4	C5	C6	C7
	House No.	Market Price	Size of House	Age of House	FITS	RESI	
1	1	63.0	1.605	35	62.466	0.5340	
2	2	65.1	2.489	45	71.465	-6.3653	
3	3	69.9	1.553	20	71.540	-1.6399	
4	4	76.8	2.404	32	78.622	-1.8218	
5	5	73.9	1.884	25	74.073	-0.1728	
6	6	77.9	1.558	14	75.627	2.2734	
7	7	74.9	1.748	8	82.991	-8.0911	
8	8	78.0	3.105	10	105.702	-27.7018	
9	9	79.0	1.682	28	68.495	10.5053	
10	10	83.4	2.470	30	81.124	2.2761	
11	11	79.5	1.820	2	88.265	-8.7649	
12	12	83.9	2.143	6	91.322	-7.4224	
13	13	79.7	2.121	14	85.602	-5.9018	
14	14	84.5	2.485	9	95.383	-10.8829	
15	15	96.0	2.300	19	85.442	10.5584	
16	16	109.5	2.714	4	102.772	6.7279	
17	17	102.5	2.463	5	97.659	4.8415	
18	18	121.0	3.076	7	107.187	13.8130	
19	19	104.9	3.048	3	109.356	-4.4563	
20	20	128.0	3.267	6	111.237	16.7625	
21	21	129.0	3.069	10	105.064	23.9361	
22	22	117.9	4.765	11	134.447	-16.5474	
23	23	140.0	4.540	8	132.460	7.5402	