

Statistics and Data Analysis

Project 2019

You have been hired by a local estate agency (Rob and Roger Estate Agents) to act as their “computer person” for your work experience. Your duties include networking, report generation, data analysis, and to provide technical support and ECDL (European Computer Driving License) training for an office with eight other employees.

Roger, the senior partner in the agency, has given you the following data set on apartment sales in the Limerick region, and requires you to perform a regression analysis of the data – more precise instructions follow. You are required to present your results in the form of a single report, having used Minitab to analyse the data.

Roger believes that it is possible to accurately estimate the price of an apartment in the Limerick region based on the following factors: age, area, number of bedrooms, provision of covered parking, floor number of main entrance, and distance from city centre. To test this belief, he has randomly selected 25 apartments and provided you with the following data set:

Selling Price (Y)	X ₁	X ₂	X ₃	X ₄	X ₅	X ₆	X ₇
€137,000	0	880	0	0	0	0	5
€141,000	1	1000	0	1	0	1	6
€158,000	0	970	0	1	1	0	13
€91,000	16	900	1	0	0	4	0
€98,000	4	830	0	1	0	0	2
€173,000	8	1100	1	0	0	2	1
€141,000	0	970	0	0	1	0	3
€119,000	11	940	0	0	1	1	0
€152,000	3	890	0	0	1	0	6
€154,000	5	1090	1	0	0	0	2
€81,000	8	730	0	0	1	3	0
€167,000	8	1200	0	1	0	6	1
€147,000	7	1080	0	0	1	2	1
€162,000	1	1100	1	0	1	0	7
€109,000	3	760	0	0	0	0	4
€158,000	1	1000	1	0	0	1	3
€153,000	2	1120	0	1	0	0	5
€151,000	7	1120	0	1	0	2	2
€195,000	7	1410	1	0	1	1	3
€122,000	0	830	0	0	1	2	7
€144,000	0	880	0	0	0	0	6
€160,000	19	1460	0	1	1	0	0
€135,000	12	1030	1	0	1	6	0
€122,000	3	770	0	0	1	4	1
€133,000	1	910	1	0	0	0	3

The predictor variables are as follows:

- X_1 = age (in years);
- X_2 = area (in square feet);
- X_3 = 1 if two bedroom, 0 otherwise;
- X_4 = 1 if three bedroom, 0 otherwise;
- X_5 = 1 if covered parking, 0 otherwise;
- X_6 = the floor number for main entrance;
- X_7 = kilometres to centre of town.

REQUIREMENTS

(1)

Fit the following regression model to the data:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \beta_5 X_5 + \beta_6 X_6 + \beta_7 X_7 + \varepsilon$$

(2)

Then use backward regression to remove any unnecessary terms step by step, summarizing the steps taken in a table of the following format (take the level of significance to be 5%).

Step	Term removed	R^2	Reason
0	n/a	???	n/a
1	???	???	???
2	???	???	???
...	...	...	...

(3)

Upon finding the best model, state clearly what this model is and explain in plain English what the estimated parameters mean.

(4)

Produce scatterplots (each with a fitted regression line) of selling price versus any predictors left in your model. In one line, explain each predictor's relationship to selling price, based on the corresponding scatterplot.

(5)

Explain how much of the variation in apartment selling price is explained by your model. Explain why you would/would not recommend using the model.

Appendix A

In Appendix A of your report, perform a residual analysis of your final model (produce the residual plots and explain whether any of the model assumptions have been violated).

Appendix B

In Appendix B of your report, show the regression analysis output generated by Minitab for the entire project, as displayed in the Session window.