

House Price Case (Regression)

Introduction

A British real estate agency will set up in the Toulouse region. It wants to know what are the most important criteria for the price of a house in order to be able to quickly assess a property.

Variables	Variable Description	Categories/levels
ID	House code	
Town	Town where the house is located	lacroix falgarde/launaguet/mons montrabe/plaisance/st jean/toulouse tournefeuille/vieille toulouse
HouseSize	Size of the house (m ²)	
NumberOfRooms	Number of rooms	
GardenSize	Size of the garden	
LivingRoomSize	Size of the living room	
DistanceFromToulouseCenter	Distance from the center of the city (km)	
AirConditioned	Is the house air conditioned ?	yes / no
Heating	Type of heating	gas/electric/solar
Price	House price (k€)	

Questions

1. Check the structure of the variables: are their type correct?
2. Study correlation among quantitative variables
3. Model the price using a multiple linear regression model
 - a. Which predictors do you choose?
 - b. Is the ANOVA test significant?
 - c. Which variables are significant in the model? Use the backward method.
 - d. Is a house located in the center of the town more expensive on average?
 - e. Are the houses with air conditioned more expensive?
 - f. Does solar heating increase the price of the house?
 - g. Evaluate the quality of the model
 - h. Study the residuals of the selected model
 - i. Is residuals means 0?
 - ii. Are residuals normally distributed?
 - iii. Are residuals homoskedastic?
 - iv. Are residuals autocorrelated?
 - v. For hypotheses not validated, use the adequate remedies.
 - i. Check for multicollinearity
 - j. What should be the price of houses with the following characteristics :

HouseSize	NumberOfRooms	GardenSize	LivingRoomSize	DistanceFromToulouseCenter	AirConditioned	Heating
150	5	700	35	10	yes	solar
110	5	80	25	0	no	electric

