Predicting Residential Home Sales Prices Using Regression Analysis

Rumil Legaspi, Rumil.legaspi@gmail.com Nancy Huerta, Email

9 April 2021

Contents

Purpose

We are conducting a multiple linear regression model from the Real Estate Sales (APPENC07) dataset to analyze the relationship of the given features, bedrooms, bathrooms, and garage size, with the outcome variable, house sales price.

Our Data

Background on dataset and variables

Our dataset is comprised of 522 total transactions from home sales during the year 2002.

Response Variable (Y)	Explanatory Variable 1 (X_1)	Explanatory Variable 2 (X_2)	Explanatory Variable 3 (X_3)
"house_price" = sales price of residence (in dollars)	"beds" = Number of bedrooms	"baths" = Number of bathrooms	"garage_size" = Number of cars the garage can hold

```
#Setting up our work environment
setwd("C:/Users/RUMIL/Desktop/APU/STAT 511 - Millie Mao (Applied Regression Analysis)/Project 2")

library(nortest)
library(olsrr)
library(car)
library(lmtest)
library(MASS)
library(tidyverse)
library(dplyr)
```

```r
#Loading in the text data
raw_data = read.table(file = "APPENC07.txt", header = FALSE, sep = "")
#Converting into tibble data frame for easier data analysis
house_data <- as_tibble(raw_data)
```

```r
#Defining and renaming our Explanatory(X) and Response(Y) variables
house_data <- house_data %>% select(house_price = V2,
                                    beds = V4,
                                    baths = V5,
                                    garage_size = V7)

#Setting explanatory and response variables
house_price <-  house_data %>% select(house_price) #Y
beds <- house_data %>% select(beds) #X1
baths <- house_data %>% select(baths) #X2
garage_size <- house_data %>% select(garage_size) #X3
```