Real Estate Data

Y = Selling Price in Tk.000 (Thousand), **X1** = Number of bedrooms, **X2** = Size of the home in square feet, **X3** = Pool (1 = yes, 0 = no), **X4** = Distance from the center of the city, **X5** = Township (1 = Gulshan, 2 = Uttara 3 = DOHS, 4 = Dhanmondi, 5 = Banani), **X6** = Garage attached (1 = yes, 0 = no), **X7** = Number of bathrooms.

S.N.	Y	X1	X2	X3	X4	X5	X6	X7
1	2630.3	4	2300	0	17	5	1	2
2	1820.4	4	2100	1	19	4	0	2
3	2420.1	3	2300	1	12	3	0	2
4	2130.6	2	2200	1	16	2	0	3
5	1390.9	2	2100	1	28	1	0	2
6	2450.4	2	2100	0	12	1	1	2
7	3270.2	6	2500	1	15	3	1	2
8	2710.8	2	2100	1	9	2	1	3
9	2210.1	3	2300	0	18	1	0	2
10	2660.6	4	2400	1	13	4	1	2
11	2920.4	4	2100	1	14	3	1	2
12	2009.0	2	1700	1	8	4	1	2
13	2700.8	6	2500	1	7	4	1	2
14	2460.1	4	2100	1	18	3	1	2
15	1940.4	2	2300	1	11	3	0	2
16	2810.3	3	2100	1	16	2	1	2
17	1720.7	4	2200	0	16	3	0	2
18	2070.5	5	2300	0	21	4	0	3
19	1980.9	3	2200	0	10	4	1	2
20	2090.3	6	1900	0	15	4	1	2
21	2520.3	4	2600	1	8	4	1	2
22	1920.9	4	1900	0	14	2	1	3
23	2090.3	5	2100	1	20	5	0	2
24	3450.3	8	2600	1	9	4	1	2
25	3260.3	6	2100	1	11	5	1	3
26	1730.1	2	2200	0	21	5	1	2
27	1807.0	2	1900	1	26	4	0	2
28	2570.2	2	2100	1	9	4	1	2
29	2303.0	3	2200	1	14	3	1	2
30	1800.4	2	2000	1	11	5	0	2
31	2304.0	2	1700	1	19	3	1	2
32	2070.1	2	2000	1	11	5	1	2
33	2470.7	5	2400	1	16	2	1	2
34	1660.2	3	2000	0	16	2	1	2
35	1770.1	2	1900	1	10	5	1	2
36	1820.7	4	2000	0	14	4	0	3
37	2160.0	4	2300	1	19	2	0	2
38	3120.1	6	2600	1	7	5	1	3
39	1990.8	3	2100	1	19	3	1	2
40	2730.2	5	2200	1	16	2	1	3
41	2060.0	3	2100	0	9	3	0	2
42	2320.2	3	1900	0	16	1	1	2
43	1980.3	4	2100	0	19	1	1	2
44	2050.1	3	2000	0	20	4	0	2
45	1750.6	4	2300	0	24	4	1	2
46	3070.8	3	2400	0	21	2	1	3
47	2690.2	5	2200	1	8	5	1	3
48	2240.8	3	2200	1	17	1	1	3
49	1710.6	3	2000	0	16	4	0	2
50	2160.8	3	2200	1	15	1	1	2
51	1920.6	6	2200	0	14	1	0	2

Dr. Md. Siddikur Rahman (Associate Professor, Dept. of Statistics, BRUR) siddikur@brur.ac.bd

52	2360.4	5	2200	1	20	3	1	2
53	1720.4	3	2200	1	23	3	0	2
54	2510.4	3	1900	1	12	2	1	2
55	2460.0	6	2300	1	7	3	1	3
56	1470.4	6	1700	0	12	1	0	2
57	1760.0	4	2200	1	15	1	1	2
58	2280.4	3	2300	1	17	5	1	2
59	1660.5	3	1600	0	19	3	0	3
60	1890.4	4	2200	1	24	1	1	2
61	3120.1	7	2400	1	13	3	1	3
62	2890.8	6	2000	1	21	3	1	3
63	2690.9	5	2200	0	11	4	1	3
64	1540.3	2	2000	1	13	2	0	2
65	2220.1	2	2100	1	9	5	1	2
66	2090.7	5	2200	0	13	2	1	2
67	1900.9	3	2200	0	18	3	1	2
68	2540.3	4	2500	0	15	3	1	2
69	2070.5	3	2100	0	10	2	0	2
70	2090.7	4	2200	0	19	2	1	2
71	2940.0	2	2100	1	13	2	1	3
72	1760.3	2	2000	0	17	3	0	2
73	2940.3	7	2400	1	8	4	1	2
74	2240.0	3	1900	0	6	1	1	2
75	1250.0	2	1900	1	18	4	0	2
76	2360.8	4	2600	0	17	5	1	2
77	1640.1	4	2300	1	19	4	0	2
78	2170.8	3	2500	1	12	3	0	2
79	1920.2	2	2400	1	16	2	0	3
80	1250.9	2	2400	1	28	1	0	2
81	2200.9	2	2300	0	12	1	1	2
82	2940.5	6	2700	1	15	3	1	2
83	2440.6	2	2300	1	9	2	1	3
84	1990.0	3	2500	0	18	1	0	2
85	2400.0	4	2600	1	13	4	1	2
86	2630.2	4	2300	1	14	3	1	2
87	1880.1	2	1900	1	8	4	1	2
88	2430.7	6	2700	1	7	4	1	2
89	2210.5	4	2300	1	18	3	1	2
90	1705.0	2	2500	1	11	3	0	2
91	2530.2	3	2300	1	16	2	1	2
92	1550.4	4	2400	0	16	3	0	2
93	1860.7	5	2500	0	21	4	0	3
94	1709.0	3	2400	0	10	4	1	2
95	1880.3	6	2100	0	15	4	1	2
96	2270.1	4	2900	1	8	4	1	2
97	1730.6	4	2100	0	14	2	1	3
98	1880.3	5	2300	1	20	5	0	2
99	3100.8	8	2900	1	9	4	1	2
100	2930.7	6	2400	1	11	5	1	3
101	1709.0	3	2400	1	8	4	1	2
102	1880.3	6	2100	0	14	2	1	3
103	2270.1	4	2900	1	20	5	0	2
104	1730.6	4	2100	1	9	4	1	2
105	1880.3	5	2300	1	11	5	1	3

1. First enter the Real estate data into R and answer the exercises.

i) Compute the Mean, Median, Mode, Standard Deviation, Variance, Quartiles, Decile, Percentile and Range of "Selling price" from the data of your sample and interpret.
ii) Develop a histogram for the variable "selling price".
iii) Develop a Pie chart and a Bar diagram for the variable "Township".
iv) Develop a Box plot for the variable "Distance".
v) What information can you give from these plots?

2. Refer to the Real Estate data, which reports information on the homes sold in Dhaka, Bangladesh, last year. Use the selling price as the dependent variable and determine the regression equation with number of bedrooms, size of the house, whether there is a pool, whether there is an attached garage, distance from the center of the city and number of bathrooms as independent variables.

i) Write out the regression equation. Discuss each of the variables. How much does a garage or a swimming pool add to the selling price of a home?
ii) Determine the value of R-square. Interpret.
iii) Determine the multiple Standard error of the estimate and interpret.
iv) Develop a correlation matrix. Which independent variables have weak or strong correlations with the dependent variable?
v) Conduct a global test on the set of independent variables. Interpret.
vi) Conduct a test of the hypothesis on each of the independent variables. Would you consider deleting any of the variables? If so, which one?
vii) Write down the new multiple regression equation.

R Codes

```
#Setting home directory
getwd()
setwd("/Users/siddikurrahman/Desktop/Economics_BRUR")
list.files()

## Problem with source
.libPaths()
.libPaths("/Users/siddikurrahman/Desktop")

#Data read in Excel
# Install packages
install.packages("readxl")
library(readxl)
# read the first worksheet from the workbook myexcel.xlsx
# first row contains variable names
dataEXCEL <- read_excel("sales.xlsx",1)
dfEXCEL <- as.data.frame(dataEXCEL)
```

```
# read in the worksheet named sheet1
dataEXCEL <- read_excel("sales.xlsx",sheet = 1)
dfEXCEL <- as.data.frame(dataEXCEL)

# read data in csv format
dataCOMMA <-read.csv("sales.csv", header=TRUE)
dataCOMMA

# Read data in SPSS
# Method 1: Using haven Package
install.packages('haven')
library(haven)
dataspss <- read_sav("real_estate.sav")

# Method 2: Using foreign Package
install.packages('foreign')
library("foreign")
dataspss1 <- read.spss("airline_passengers.sav",
to.data.frame = TRUE)
head(dataspss1)

# Method 3: Using Hmisc Package
install.packages('Hmisc')
library("Hmisc")
data2 <- spss.get("airline_passengers.sav",
to.data.frame = TRUE)
head(data2)

# file.choose() function
datacsv<-read.csv(file.choose(),header = T);d
dataEXCELL <- read_excel(file.choose())
dataspss=read_sav(file.choose())

# Show data
head(dataspss)
dataspss <- as.data.frame(dataspss)
dataspss
names(dataspss)

# check if necessary/not:
options(max.print=999999)

# Check missing value
any(is.na(dataspss))
head(dataspss)
```

```
# Descriptive statistics
summary(dataspss)

# Basic Graphics
Selling_price=dataspss$Y
Number_of_bedroom=dataspss$X1
Home_size=dataspss$X2
Pool=dataspss$X3
Distance_from_the_center_of_the_city=dataspss$X4
Township = factor(dataspss$X5,  labels = c("Gulshan", "Uttara", "DOHS", "Dhanmondi", "Banani"))
Garage_attached=dataspss$X6
Number_of_bathrooms=dataspss$X7
x=table(Township)
x
#adjust plot margins: Figures margin too large
par(mar = c(1, 1, 1, 1))
# Bar plot
barplot(x, main="Barplot of Township",col = "red")

#Histogram
hist(dataspss$Y,main= "Histogram of Selling Price", col = "green")
#Pie Chart
pie(x,main="Pie Chart of Township")

#Bivariate analysis:scatter plot with x and y
plot(dataspss$X1, dataspss$Y, main = "Scatter plot of X1 and Y")

#Outlier handling:boxplot of x and y
box = boxplot(dataspss$Y)

# Correlation in R
# cor Function
cor(dataspss)
#The rcorr Function
#create matrix of correlation coefficients and p-values
library(Hmisc)
rcorr(as.matrix(dataspss))

#The corrplot Function
install.packages("corrplot")
library(corrplot)
corrplot(cor(dataspss))

#visualize correlation matrix
install.packages("ggplot")
install.packages("ggcorrplot")
library(ggcorrplot)
ggcorrplot(cor(dataspss))
```

 # Regression in R

```
lmrealestate = lm(Y~X1, data = dataspss) #Create the linear regression
summary(lmrealestate) #Review the results
lmrealestate = lm(Y~X1+X2+X3+X4+X5+X6+X7, data = dataspss) #Create the linear regression
summary(lmrealestate) #Review the results
```

 # Prediction Regression in R

```
dataspss$predicted = predict(lmrealestate)
dataspss$predicted
```