Build a model predicting Value

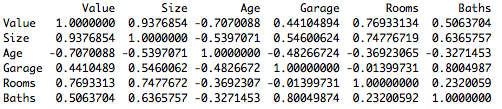
1. Comment on R2 and adjusted R2. Interpret adjusted R2.

R2 is 0.9599. This means that 95.99% of variation is explained by the regression model. Adjusted R2 is .8931, which means that 89.31% of the variation is explained by the regression, but it is adjusted for the number of X or predictor variables (degrees of freedom). This is still a high number, but it is 6% lower than R2, which indicates that the x variables are not providing a good benefit in predicting the Y.

1. Build a scatterplot matric and comment appropriately.

Value and size are fairly correlated with a positive slope. The pair Size and Age show some signs of correlation, while all other pairs seem to be uncorrelated. 

1. Build a correlation matrix and comment.



Value and size are the most closely correlated with 0.937685, so the strongest correlation in predicting value is size.

1. Interpret the slope for Garage.

The slope for Garage is -38.4491. This means that all variables aside, each additional garage is a decrease of $38,449 in value of house.

1. Which variables are significant? How do you know?

None of the variables are significant, because none of the Pr(>|t|) values in the model are below 0.05, meaning they are not significant.

CODE

homes <- read.csv(file.choose(), header=TRUE)

attach(homes)

model1 = lm(Value ~ Size + Age + Garage + Rooms + Baths)

summary(model1)

//

Call:

lm(formula = Value ~ Size + Age + Garage + Rooms + Baths)

Residuals:

1 2 3 4 5 6 7 8 9

5.257 -13.719 11.236 -66.261 -3.276 7.706 2.882 50.270 5.904

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 110.1807 154.2452 0.714 0.5266

Size 0.2162 0.0800 2.703 0.0736 .

Age -2.1893 1.0160 -2.155 0.1201

Garage -38.4491 57.7353 -0.666 0.5531

Rooms -0.7231 33.4941 -0.022 0.9841

Baths 1.1132 45.3529 0.025 0.9820

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 49.57 on 3 degrees of freedom

Multiple R-squared: 0.9599, Adjusted R-squared: 0.8931

F-statistic: 14.36 on 5 and 3 DF, p-value: 0.02628

Code:

Varb = data.frame(Value, Size, Age,Garage, Rooms, Baths)

pairs(Varb, upper.panel=NULL)

cor(cbind(Value, Size, Age, Garage, Rooms, Baths))

Value Size Age Garage Rooms Baths

Value 1.0000000 0.9376854 -0.7070088 0.44104894 0.76933134 0.5063704

Size 0.9376854 1.0000000 -0.5397071 0.54600624 0.74776719 0.6365757

Age -0.7070088 -0.5397071 1.0000000 -0.48266724 -0.36923065 -0.3271453

Garage 0.4410489 0.5460062 -0.4826672 1.00000000 -0.01399731 0.8004987

Rooms 0.7693313 0.7477672 -0.3692307 -0.01399731 1.00000000 0.2320059

Baths 0.5063704 0.6365757 -0.3271453 0.80049874 0.23200592 1.0000000

> plot(Value ~ Garage)

> model2 = lm(Value ~ Garage)

> summary(model2)

Call:

lm(formula = Value ~ Garage)

Residuals:

Min 1Q Median 3Q Max

-136.25 -75.48 -38.95 20.16 310.22

Coefficients:

Estimate Std. Error t value Pr(>|t|)

(Intercept) 353.95 75.96 4.659 0.00232 \*\*

Garage 85.54 65.79 1.300 0.23470

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 145.5 on 7 degrees of freedom

Multiple R-squared: 0.1945, Adjusted R-squared: 0.07946

F-statistic: 1.691 on 1 and 7 DF, p-value: 0.2347