Stat 415 Regression: Classwork/Lab 3

Frankie Tyndall

2022-07-07

## Study of Selling Price of a Home

Study looks at the amount of taxes applied to specific selling prices of a home.

x = Taxes

y = selling prices

Selling\_home <- tribble(~X, ~Y,  
 3104, 279900,  
 1173, 146500,  
 3076, 237700,  
 1608, 200000,  
 1454, 159900,  
 2997, 499900,  
 4054, 265500,  
 3002, 289900  
 )   
  
  
Selling\_home

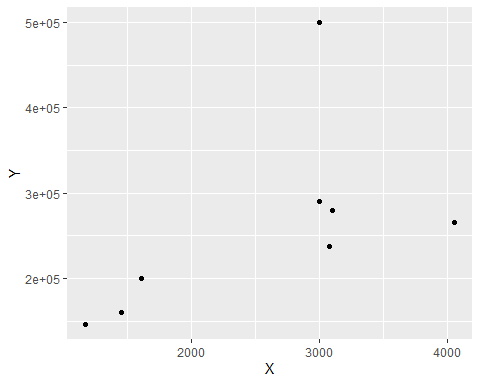
## # A tibble: 8 × 2  
## X Y  
## <dbl> <dbl>  
## 1 3104 279900  
## 2 1173 146500  
## 3 3076 237700  
## 4 1608 200000  
## 5 1454 159900  
## 6 2997 499900  
## 7 4054 265500  
## 8 3002 289900

## Hypothesis for F-Test

H0:β1=0 # null hypothesis: linear relationship does not exist HA:β1≠0 # alternative hypothesis: linear relationship does exist

## Use and show R code to produce a scatter plot for the bivariate data.Indicate if your scatter plot shows an outlier.

qplot(x = X, y = Y, data = Selling\_home, geom = "point")

 The scatterplot shows that there are one or more outliers present in the bivaridate data. Outliers negatively impact normality, but the scatterplot does suggest a positive linear relationship.

## Use and show R code to produce a linear model for the bivariate data in the table.

lm(Y~X, data = Selling\_home)

##   
## Call:  
## lm(formula = Y ~ X, data = Selling\_home)  
##   
## Coefficients:  
## (Intercept) X   
## 100456.77 62.32

slope = 62.32 y intercept = 100456.77 y(hat) = 100456.77 + 62.32X

## Use and show R code to produce all of the residuals for your model.

model8 <- lm(Y~X, data = Selling\_home)  
residual\_model <- resid(model8)  
residual\_model

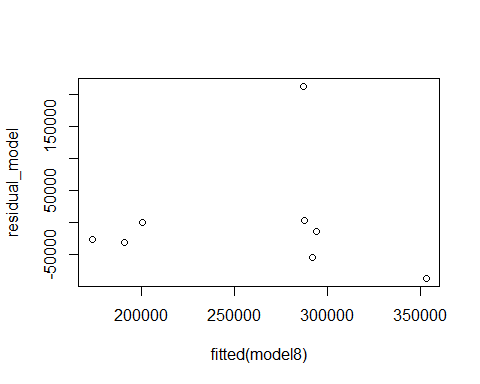
## 1 2 3 4 5 6   
## -14010.1936 -27062.7210 -54465.1241 -673.6224 -31175.7400 212658.4649   
## 7 8   
## -87617.9093 2346.8454

## Use and show R code to produce a residual plot for the data. Does your residual plot also suggest the existence of an outlier?

model8 <- lm(Y~X, data = Selling\_home)  
residual\_model <- resid(model8)  
residual\_model

## 1 2 3 4 5 6   
## -14010.1936 -27062.7210 -54465.1241 -673.6224 -31175.7400 212658.4649   
## 7 8   
## -87617.9093 2346.8454

plot(fitted(model8), residual\_model)

 The residual plot does suggest the existence of an outlier in the data. There are residuals that stand out from the basic pattern of residuals.

## Execute an F test in order to determine if a linear model is appropriate. Use the steps and procedure illustrated in class by making use of an ANOVA table ,the F value, and the F critical number. And of course, indicate if the null hypothesis should be rejected.

# 8-2 = 6 degrees of freedom   
model8 <- lm(Y~X, data = Selling\_home)  
anova(model8)

## Analysis of Variance Table  
##   
## Response: Y  
## Df Sum Sq Mean Sq F value Pr(>F)  
## X 1 2.8098e+10 2.8098e+10 2.9181 0.1385  
## Residuals 6 5.7774e+10 9.6289e+09

qf(p=.05, df1=1, df2=6, lower.tail=FALSE)

## [1] 5.987378

The F value is 2.9181 The the F critical number is 5.987378

Since the F value (2.9181) is less than the F critical number 5.987378, we fail to reject the null hypothesis that B1 = 0. This means that a linear relationship does not exist between X (Taxes) and Y(Selling Price).

## Which of the two residual plots shown above suggests that a linear model for the associated bivariate data is not appropriate? Provide a brief comment of justification for your answer.

The residual plot of B suggests that a linear model for the associated bivariate data is not appropriate because the plot is patterned and curved which are not indicators of a linear model.