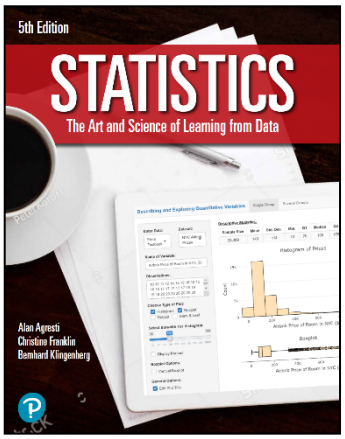

R Code for Examples in the book
"Statistics: The Art and Science of Learning from Data"
by Agresti, Franklin and Klingenberg, 5th edition

Chapter 13

Example 3: Predicting House Prices – Multiple Correlation Coefficient and R^2

Reading in data

```
houses <-
read.csv(file='https://raw.githubusercontent.com/artofstat/data/master/Chapter13/house_selling_prices_or.csv')
```

Fitting in multiple regression model

```
linReg <- lm(HP.in.thousands ~ House.Size + Bedrooms, data = houses)
linReg

##
## Call:
## lm(formula = HP.in.thousands ~ House.Size + Bedrooms, data = houses)
##
## Coefficients:
## (Intercept)   House.Size     Bedrooms
##    60.10214      0.06298     15.17041
```

To get the ANOVA table for the regression model

```
aov <- anova(linReg)
aov

## Analysis of Variance Table
##
## Response: HP.in.thousands
##             Df  Sum Sq Mean Sq  F value    Pr(>F)
## House.Size   1 1347323 1347323 209.1019 < 2.2e-16 ***
## Bedrooms     1   52202   52202   8.1016  0.004891 **
## Residuals  197 1269345    6443
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

To compute R squared using sum of squares

```
tss <- sum(aov$`Sum Sq`)
rss <- aov$`Sum Sq`[3]
rSquared <- (tss - rss) / tss
rSquared
```

```
## [1] 0.5243884
```

To find the multiple correlation coefficient

```
r <- sqrt(rSquared)
r
```

```
## [1] 0.7241467
```

To verify that the output for R Squared is correct using the manual computation, you can use the `summary()` function on our model; the R squared is shown there as well

```
summary(linReg)
```

```
##
## Call:
## lm(formula = HP.in.thousands ~ House.Size + Bedrooms, data = houses)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -306.92  -35.16   -0.75   30.47  376.81
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 60.102140  18.622905   3.227  0.00146 **
## House.Size   0.062983   0.004753  13.250  < 2e-16 ***
## Bedrooms    15.170411   5.329806   2.846  0.00489 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 80.27 on 197 degrees of freedom
## Multiple R-squared:  0.5244, Adjusted R-squared:  0.5196
## F-statistic: 108.6 on 2 and 197 DF,  p-value: < 2.2e-16
```