MLR Example

SDS 291

2/19/2020

We're using data from a sample of 104 homes in Northampton, MA to see whether being close to the bike trail enhances the value of the home. Specifically, we're looking at the association between square feet (a house's size) and distance from the rail trail with the house's estimated value in 2014. The variables we're using are:

- `Price2014`: Zillow price estimate from 2014 (in thousands of dollars)
- `Distance`: Distance (in miles) to the nearest entry point to the rail trail network
- `SquareFeet`: Square footage of interior finished space (in thousands of sf)

Regression Output

```
m1<-lm(Price2014 ~ SquareFeet + Distance , data = RailsTrails)
summary(m1)
```

```
##
## Call:
## lm(formula = Price2014 ~ SquareFeet + Distance, data = RailsTrails)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -152.15  -30.27   -4.14   25.75  337.93
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   78.985     25.607   3.085  0.00263 **
## SquareFeet   147.920     12.765  11.588  < 2e-16 ***
## Distance     -15.788      7.586  -2.081  0.03994 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 65.55 on 101 degrees of freedom
## Multiple R-squared:  0.6574, Adjusted R-squared:  0.6506
## F-statistic: 96.89 on 2 and 101 DF,  p-value: < 2.2e-16
```

1. Write the fitted regression equation.

2. Test the hypothesis that distance from the rail trail has a linear relationship with house price in 2014.

3. Calculate the 95% confidence interval for Distance to 3 decimal places (t* = 1.96) and interpret in a sentence.

4. What price would this model predict for a 1700 square foot house that it .986 miles from the rail trail? (Be cautious with the units)

4. An actual house in this dataset that is 1700 square feet and .986 miles from the rail trail entrance had an Zillow price estimate of $222,864. Calculate the residual for this house and interpret it in a sentence in the context of this problem.

Residuals and Model Error

```
anova(m1)
```

```
## Analysis of Variance Table
##
## Response: Price2014
##             Df Sum Sq Mean Sq  F value  Pr(>F)
## SquareFeet   1 813976  813976 189.4454 < 2e-16 ***
## Distance     1  18611   18611   4.3316 0.03994 *
## Residuals  101 433959    4297
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

1. Calculate the R^2 (same equation as simple linear regression, p.103-104) and Adjusted R^2 for the model (see p.105 for the equation). Interpret each in a sentence.

2. Calculate the regression standard error for this model (p.99) and interepret it in a sentence.

3. Calculate the F statistic for the model: $F = \frac{MSModel}{MSError}$ (see p.102 for making sense of the ANOVA table)

4. State the null and alternative hypotheses for the F test (see p.102). Look at the F Distribution calculator (at https://gallery.shinyapps.io/dist_calc/) and estimate the p-value for your F statistic with the degrees of freedom above (it will be an approximation - the slider bars won't go as high as you need them to - but it gives you the rough answer for your hypothesis). What do you conclude about your hypothesis?