

# MLR Examples

SDS 291

2/24/2020

## Rail Trail Multiple Regression Example

We're still 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)

```
library(Stat2Data)
data("RailsTrails")
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. What price would this model predict for a 1000 square foot house that is *1 mile* from the rail trail? (Be cautious with the units)

1. What price would this model predict for a 1000 square foot house that is *2 miles* from the rail trail?  
(Be cautious with the units)

## Adjusting for Distance Group

Rather than distance in miles, what if we thought a more useful measure would be whether the house was closer ( $< 1$  mile from an entrance to the rail trail) or further away ( $\geq 1$  mile from a rail trail entrance)?

- **Price2014**: Zillow price estimate from 2014 (in thousands of dollars)
- **DistGroup**:
  - **Closer**:  $< 1$  mile to the nearest entry point to the rail trail network
  - **Farther Away**:  $\geq 1$  mile to the nearest entry point to the rail trail network
- **SquareFeet**: Square footage of interior finished space (in thousands of sf)

R treats **DistGroup** as a **factor** variable. It can also be treated as a numeric variable, where one category has the value of 0 and the other category has the value of 1. In other words, you can think that the numerical equivalent is: “Closer” = 0 and “Farther Away” = 1.

```
m2<-lm(Price2014 ~ SquareFeet+DistGroup , data = RailsTrails)
summary(m2)
```

```
##
## Call:
## lm(formula = Price2014 ~ SquareFeet + DistGroup, data = RailsTrails)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -136.55  -30.14   -2.14    22.17   321.40
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)       80.10      23.13   3.463 0.000785 ***
## SquareFeet       150.50      11.83  12.724 < 2e-16 ***
## DistGroupFarther Away  -36.97      13.51  -2.736 0.007356 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 64.59 on 101 degrees of freedom
## Multiple R-squared:  0.6673, Adjusted R-squared:  0.6607
## F-statistic: 101.3 on 2 and 101 DF,  p-value: < 2.2e-16
```

1. What price would this model predict for a 1000 square foot house that is *Closer* from the rail trail?

1. What price would this model predict for a 1000 square foot house that is *Farther Away* from the rail trail?

1. What price would this model predict for a 2000 square foot house that is *Closer* from the rail trail?

1. What price would this model predict for a 2000 square foot house that is *Farther Away* from the rail trail?

## Bedrooms

Rather than square feet, let's consider the number of bedrooms the house has, in addition to its distance from the rail trail.

- **Price2014**: Zillow price estimate from 2014 (in thousands of dollars)
- **BedGroup**: Categorical Variable of house type by group of bedrooms:
  - 1-2 bedrooms (reference),
  - 3 bedrooms,
  - 4+ bedrooms
- **Distance**: Distance (in miles) to the nearest entry point to the rail trail network

You can think about **BedGroup** similarly to **DistGroup** and consider the 3 bedroom group output in the model below akin to an indicator variable with the values of 0 or 1: 0 if the house doesn't have 3 bedrooms and 1 if it does have 3 bedrooms. Same for 4+ bedrooms.

```
m3 <- lm(Price2014 ~ Distance+BedGroup, data = RailsTrails)
summary(m3)
```

```
##
## Call:
## lm(formula = Price2014 ~ Distance + BedGroup, data = RailsTrails)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -195.28  -48.15  -13.19   26.02  509.02
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      283.56      26.67  10.633 < 2e-16 ***
## Distance         -42.30      10.18  -4.154 6.89e-05 ***
## BedGroup3 beds    39.88      26.63   1.498 0.137364
## BedGroup4+ beds  106.13      28.49   3.725 0.000323 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 93.12 on 100 degrees of freedom
## Multiple R-squared:  0.3153, Adjusted R-squared:  0.2948
## F-statistic: 15.35 on 3 and 100 DF,  p-value: 2.746e-08
```

## 1000 Square Feet Houses

1. What price would this model predict for a 1000 square foot house that has 1-2 Bedrooms?

1. What price would this model predict for a 1000 square foot house that has 3 bedrooms?

1. What price would this model predict for a 1000 square foot house that has 4+ bedrooms?

## **2000 Square Feet Houses**

1. What price would this model predict for a 2000 square foot house that has 1-2 Bedrooms?

1. What price would this model predict for a 2000 square foot house that has 3 bedrooms?

1. What price would this model predict for a 2000 square foot house that has 4+ bedrooms?

## Back to Distance Group

What if we thought that the Square Footage of a house would *vary* by whether it's closer or further from the rail trail. A big house may not matter as much if it's really far from the rail trail, and a smaller house may be more valuable if it's closer to the rail trail than if it were further away.

```
m4 <- lm(Price2014 ~ SquareFeet*DistGroup , data = RailsTrails)
summary(m4)

##
## Call:
## lm(formula = Price2014 ~ SquareFeet * DistGroup, data = RailsTrails)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -130.287  -32.792    0.084   23.018  282.596
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)         32.16      36.11   0.891  0.3752
## SquareFeet         177.82      19.75   9.003 1.51e-14 ***
## DistGroupFarther Away    32.46      42.57   0.763  0.4475
## SquareFeet:DistGroupFarther Away -42.15      24.53  -1.718  0.0889 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 63.97 on 100 degrees of freedom
## Multiple R-squared:  0.6769, Adjusted R-squared:  0.6672
## F-statistic: 69.82 on 3 and 100 DF,  p-value: < 2.2e-16
##
## #qplot(y=Price2014, x=SquareFeet, data=RailsTrails, color=DistGroup) +
## # geom_smooth(method=lm,se=FALSE,fullrange = TRUE)
```

1. What price would this model predict for a 1000 square foot house that is *Closer* from the rail trail?

1. What price would this model predict for a 1000 square foot house that is *Farther Away* from the rail trail?

1. What price would this model predict for a 2000 square foot house that is *Closer* from the rail trail?

1. What price would this model predict for a 2000 square foot house that is *Farther Away* from the rail trail?