Interaction Term Testing

Tarek El-Hajjaoui

2023-02-09

Loading the dataset

```
file_path = '/Users/Tarek/Documents/UCI_MDS_Coding/Stats210P/R_Statistical_Modeling/Datasets/MidwestSal
df = read.table(file_path, header=FALSE, sep="", dec=".")
```

Give labels to the columns of the dataset.

```
names(df)=c("id","price","sqft","bed","bath","ac","garage","pool","year","quality","style","lot","hwy")
```

Summary of data set

```
str(df)
```

```
## 'data.frame':    522 obs. of  13 variables:
##  $ id     : int  1 2 3 4 5 6 7 8 9 10 ...
##  $ price  : int  360000 340000 250000 205500 275500 248000 229900 150000 195000 160000 ...
##  $ sqft   : int  3032 2058 1780 1638 2196 1966 2216 1597 1622 1976 ...
##  $ bed    : int  4 4 4 4 4 4 3 2 3 3 ...
##  $ bath   : int  4 2 3 2 3 3 2 1 2 3 ...
##  $ ac     : int  1 1 1 1 1 1 1 1 1 0 ...
##  $ garage : int  2 2 2 2 2 5 2 1 2 1 ...
##  $ pool   : int  0 0 0 0 0 1 0 0 0 0 ...
##  $ year   : int  1972 1976 1980 1963 1968 1972 1972 1955 1975 1918 ...
##  $ quality: int  2 2 2 2 2 2 2 2 3 3 ...
##  $ style  : int  1 1 1 1 7 1 7 1 1 1 ...
##  $ lot    : int  22221 22912 21345 17342 21786 18902 18639 22112 14321 32358 ...
##  $ hwy    : int  0 0 0 0 0 0 0 0 0 0 ...
```

Creating 2 different models with the same predictors except Model 2 will have an added interaction term.

$$\hat{Y} = \hat{Price}$$
$$X_1 = sqft$$
$$X_2 = NumOfBedroomrs$$
$$X_1 * X_2 = sqft * NumOfBedroomrs$$

Model 1

$$\hat{Y} = \hat{\beta_0} + \hat{\beta_1} X_1 + \hat{\beta_2} X_2$$

```
# Model 1
model_1 = lm(price ~ sqft + bed, data=df)
summary(model_1)
```

```
##
## Call:
## lm(formula = price ~ sqft + bed, data = df)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -227961  -38270   -8693   24670  381949
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -66971.563  13404.170  -4.996    8e-07 ***
## sqft           165.832      5.855  28.326   <2e-16 ***
## bed          -8647.511   4104.008  -2.107   0.0356 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 78860 on 519 degrees of freedom
## Multiple R-squared:  0.6743, Adjusted R-squared:  0.6731
## F-statistic: 537.3 on 2 and 519 DF,  p-value: < 2.2e-16
```

Model 2

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2 + \hat{\beta}_3 (X_1 * X_2)$$

```
# Model 2
model_2 = lm(price ~ sqft + bed + sqft * bed, data=df)
summary(model_2)
```

```
##
## Call:
## lm(formula = price ~ sqft + bed + sqft * bed, data = df)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -196725  -41800   -6185   27098  376624
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -2.033e+05  3.625e+04  -5.607 3.36e-08 ***
## sqft         2.281e+02  1.646e+01  13.857  < 2e-16 ***
## bed          2.754e+04  9.835e+03   2.800  0.00529 **
## sqft:bed    -1.576e+01  3.904e+00  -4.037 6.23e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 77730 on 518 degrees of freedom
## Multiple R-squared:  0.6843, Adjusted R-squared:  0.6824
## F-statistic: 374.2 on 3 and 518 DF,  p-value: < 2.2e-16
```