

Midwest Sales

Tarek El-Hajjaoui

2023-01-29

Loading the dataset

```
file_path = '/Users/Tarek/Documents/UCI_MDS_Coding/Stats210P/R_Statistical_Modeling/MidwestSales/MidwestSales.csv'
Midwest = read.table(file_path, header=FALSE, sep=" ", dec=".")
```

Give labels to the columns of the dataset.

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

```
##      id      price      sqft      bed
## Min.   : 1.0    Min.   : 84000  Min.   : 980  Min.   :0.000
## 1st Qu.:131.2    1st Qu.:180000  1st Qu.:1701  1st Qu.:3.000
## Median :261.5    Median :229900  Median :2061  Median :3.000
## Mean   :261.5    Mean   :277894  Mean   :2261  Mean   :3.471
## 3rd Qu.:391.8    3rd Qu.:335000  3rd Qu.:2636  3rd Qu.:4.000
## Max.   :522.0    Max.   :920000  Max.   :5032  Max.   :7.000
##      bath      ac      garage      pool
## Min.   :0.000    Min.   :0.00000  Min.   :0.0   Min.   :0.00000
## 1st Qu.:2.000    1st Qu.:1.0000  1st Qu.:2.0   1st Qu.:0.00000
## Median :3.000    Median :1.0000  Median :2.0   Median :0.00000
## Mean   :2.642    Mean   :0.8314  Mean   :2.1   Mean   :0.06897
## 3rd Qu.:3.000    3rd Qu.:1.0000  3rd Qu.:2.0   3rd Qu.:0.00000
## Max.   :7.000    Max.   :1.0000  Max.   :7.0   Max.   :1.00000
##      year      quality      style      lot
## Min.   :1885    Min.   :1.000    Min.   : 1.000  Min.   : 4560
## 1st Qu.:1956    1st Qu.:2.000    1st Qu.: 1.000  1st Qu.:17205
## Median :1966    Median :2.000    Median : 2.000  Median :22200
## Mean   :1967    Mean   :2.184    Mean   : 3.345  Mean   :24370
## 3rd Qu.:1981    3rd Qu.:3.000    3rd Qu.: 7.000  3rd Qu.:26787
## Max.   :1998    Max.   :3.000    Max.   :11.000  Max.   :86830
##      hwy
## Min.   :0.00000
## 1st Qu.:0.00000
## Median :0.00000
## Mean   :0.02107
## 3rd Qu.:0.00000
## Max.   :1.00000
```

Hypothesis Testing (two-sided test)

- Null Hypothesis: there is no significant linear relationship between the square footage and price of a house; ($B_1 = 0$).
- Alternative Hypothesis: there is a significant linear relationship between the square footage and price of a house; ($B_1 \neq 0$).

Confidence Intervals

Creating a Linear Model where the square footage of the house is used to predict the sale price.

```
model <- lm(price ~ sqft, data=Midwest)
summary(model)

##
## Call:
## lm(formula = price ~ sqft, data = Midwest)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -239405  -39840   -7641   23515  388362
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -81432.946  11551.846   -7.049 5.74e-12 ***
## sqft         158.950     4.875   32.605 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 79120 on 520 degrees of freedom
## Multiple R-squared:  0.6715, Adjusted R-squared:  0.6709
## F-statistic: 1063 on 1 and 520 DF,  p-value: < 2.2e-16
```

Confidence Interval (CI) for sqft B0 value

```
confint(model)

##              2.5 %      97.5 %
## (Intercept) -104126.9690 -58738.9238
## sqft         149.3731    168.5273
```

95% confidence interval for the mean value of the response (price) when sqft=2000

```
# interval="c" means confidence interval.  
predict(model, list(sqft=2000), interval= "c")
```

```
##          fit          lwr          upr  
## 1 236467.5 229220.7 243714.4
```

95% prediction interval for the mean value of the response (price) when sqft=2000

```
# interval="c" means confidence interval.  
predict(model, list(sqft=2000), interval= "p")
```

```
##          fit          lwr          upr  
## 1 236467.5 80858.85 392076.2
```

90% prediction interval for the mean value of the response (price) when sqft=2000

```
# interval="c" means confidence interval.  
predict(model, list(sqft=2000), interval= "p", level = 0.90)
```

```
##          fit          lwr          upr  
## 1 236467.5 105948 366987
```