Regressionary Difficulties: Windsor, Canada

Katelyn Patricio

2025-06-29

Modeling the impact of location on price

```
ch <- lm(price ~ lotsize + bedrooms + bathrooms + recreation + garage + prefer, data = HousePrices)

summary(ch)
```

```
##
## Call:
## lm(formula = price ~ lotsize + bedrooms + bathrooms + recreation +
##     garage + prefer, data = HousePrices)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -55993 -11115  -1030   8416  80215
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)     838.434   3530.548   0.237    0.812
## lotsize           4.063      0.391  10.390  < 2e-16 ***
## bedrooms       5099.970   1122.764   4.542 6.87e-06 ***
## bathrooms     18297.819   1670.862  10.951  < 2e-16 ***
## recreationyes  8213.116   2044.259   4.018 6.71e-05 ***
## garage         5099.346    955.442   5.337 1.39e-07 ***
## preferyes     11606.292   1869.402   6.209 1.07e-09 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 17810 on 539 degrees of freedom
## Multiple R-squared:  0.5599, Adjusted R-squared:  0.555
## F-statistic: 114.3 on 6 and 539 DF,  p-value: < 2.2e-16
```

Modeling the impact of AC on price

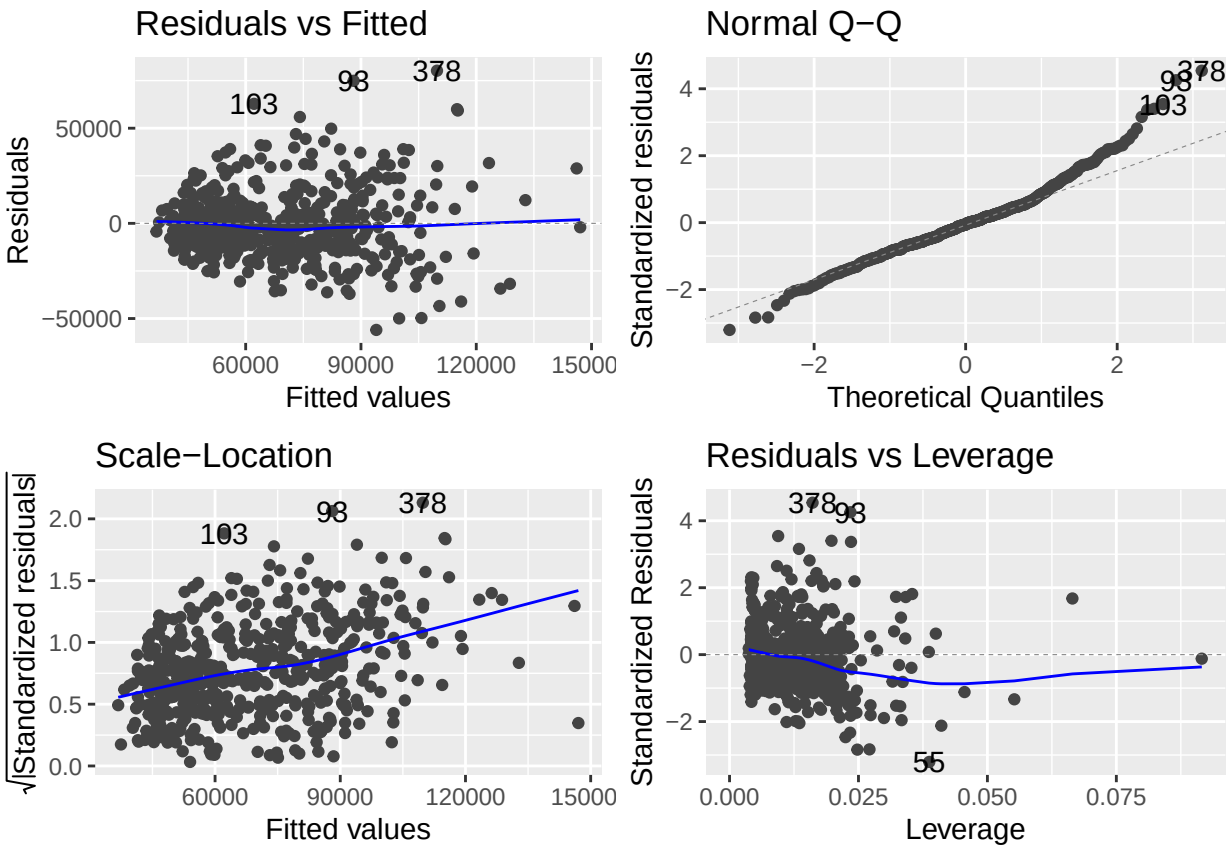
```
ac <- lm(price ~ lotsize + bedrooms + bathrooms + recreation + garage + aircon, data = HousePrices)

summary(ac)
```

```
##
```

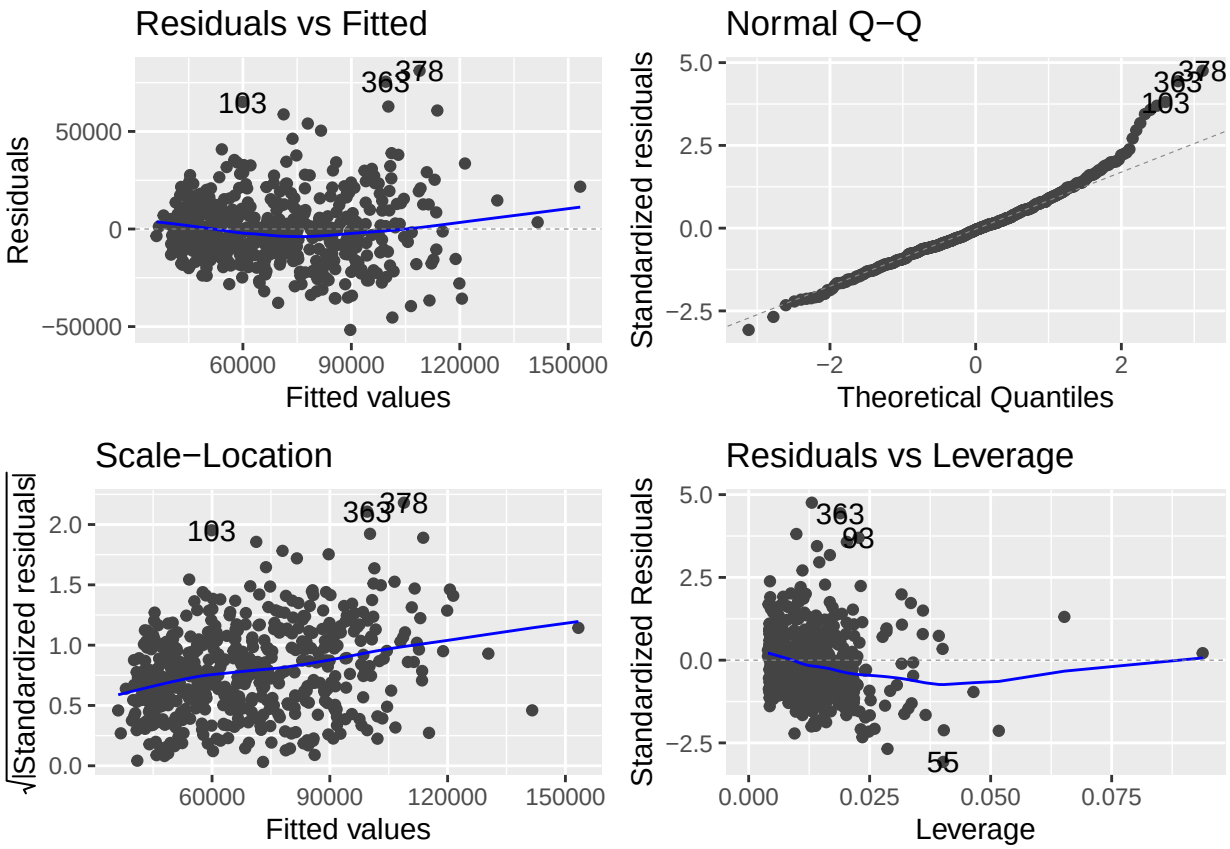
```
## Call:
## lm(formula = price ~ lotsize + bedrooms + bathrooms + recreation +
##     garage + aircon, data = HousePrices)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -51715 -10446   -873   9413  81126
##
## Coefficients:
##                Estimate Std. Error t value Pr(>|t|)
## (Intercept)   2.717e+03  3.413e+03   0.796    0.426
## lotsize       4.046e+00  3.735e-01  10.831  < 2e-16 ***
## bedrooms      4.574e+03  1.085e+03   4.217 2.91e-05 ***
## bathrooms     1.681e+04  1.618e+03  10.390  < 2e-16 ***
## recreationyes 8.171e+03  1.962e+03   4.165 3.63e-05 ***
## garage        4.586e+03  9.229e+02   4.969 9.06e-07 ***
## airconyes     1.506e+04  1.655e+03   9.101  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 17170 on 539 degrees of freedom
## Multiple R-squared:  0.5912, Adjusted R-squared:  0.5867
## F-statistic: 129.9 on 6 and 539 DF,  p-value: < 2.2e-16
```

Checking if each model is corrupted by regression violation assumptions

```
autoplot(ch)
```


Residuals vs Fitted
Residuals
Fitted values
103
98
378
Normal Q−Q
Standardized residuals
Theoretical Quantiles
Scale−Location
√|Standardized residuals|
Fitted values
Residuals vs Leverage
Standardized Residuals
Leverage
55

```
autoplot(ac)
```



Conclusion:

In both models, all variables are statistically significant (three stars) suggesting that all of these impact housing prices. Focusing on our first model, location has a large impact on housing price even after controlling for all of the other variables. Looking at its coefficent, it seems that homes in preferred areas add about $11,606 to its price. On the other hand, the same goes for aircon. A house that has AC increases the house price by approximately $15,000. Although these are indicated, there could be some outliers in this data that could be skewing the results according to our diagnostic plots. With that being said, location and aircon do have an impact on housing prices, but additional examinations might be needed to refine the model due to the outcomes of our diagnostic plots.