Modelling Housing Prices

Karen Hovhannisyan, Trisha Truong, Hao Gao

Collecting Data

Data set from Ames, Iowa Assessor’s Office

Collected data on 80 different aspects of houses

Includes: Lot Area, Year Built, Year Remodelled, Square Footage, Bedrooms, Bathrooms, etc.

As well as qualitative variables: Rating of Condition, Lot Shape, Neighborhood, House Style, Exterior Material, Foundation

Linear Regression Model

Following:

$$\hat{Y}_i = \mathbf{XB}$$

$$\text{where } \mathbf{B} = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \dots \\ \beta_{28} \end{pmatrix}, \mathbf{X} = \begin{pmatrix} X_{i1} \\ X_{i2} \\ \dots \\ X_{i3} \\ X_{i28} \end{pmatrix}$$

Output helps determine which variables to include

Gives beta values for X variables

```
Call:
lm(formula = Price ~ ., data = data_with_interaction)

Residuals:
    Min      1Q  Median      3Q     Max
-279120  -16203   -1265   14898  255605

Coefficients:
                              Estimate Std. Error t value Pr(>|t|)
(Intercept)                  1.685e+05  9.636e+02 174.846  < 2e-16 ***
LotArea                      9.831e-01  1.869e-01   5.259 1.58e-07 ***
YearBuilt                    3.461e+02  4.361e+01   7.935 3.29e-15 ***
YearRemodAdd                 5.716e+02  5.086e+01  11.239  < 2e-16 ***
MasVnrArea                   9.990e+00  5.257e+00   1.900 0.057525 .
BsmtFinSF1                   2.027e+01  2.329e+00   8.702  < 2e-16 ***
TotalBsmtSF                  4.083e+01  2.465e+00  16.567  < 2e-16 ***
BsmtFullBath                 6.039e+03  1.769e+03   3.413 0.000653 ***
FullBath                     1.982e+04  1.872e+03  10.588  < 2e-16 ***
HalfBath                     2.081e+04  1.715e+03  12.137  < 2e-16 ***
Bedroom                     -4.068e+03  1.281e+03  -3.174 0.001522 **
Kitchen                     -4.752e+04  4.427e+03 -10.734  < 2e-16 ***
TotRmsAbvGrd                 9.359e+03  7.870e+02  11.892  < 2e-16 ***
Fireplaces                   1.197e+04  1.280e+03   9.355  < 2e-16 ***
GarageCars                   6.172e+03  2.150e+03   2.871 0.004132 **
GarageArea                   4.161e+01  7.063e+00   5.892 4.40e-09 ***
WoodDeckSf                   1.388e+01  5.995e+00   2.315 0.020702 *
EnclosedPorch                2.491e+01  1.147e+01   2.172 0.029974 *
ScreenPorch                  6.125e+01  1.202e+01   5.097 3.74e-07 ***
Miscval                     -1.166e+01  1.455e+00  -8.015 1.76e-15 ***
`LotArea:YearBuilt`          3.323e-02  6.508e-03   5.106 3.56e-07 ***
`LotArea:MasVnrArea`         1.885e-03  6.922e-04   2.723 0.006525 **
`LotArea:BsmtFinSF1`        -4.167e-03  3.125e-04 -13.335  < 2e-16 ***
`LotArea:BsmtFullBath`       1.731e+00  2.684e-01   6.449 1.38e-10 ***
`LotArea:FullBath`          -1.518e+00  3.019e-01  -5.028 5.36e-07 ***
`LotArea:Bedroom`            6.983e-01  1.891e-01   3.693 0.000227 ***
`LotArea:Fireplaces`         6.018e-01  2.313e-01   2.602 0.009334 **
`LotArea:GarageCars`         1.938e+00  3.820e-01   5.073 4.24e-07 ***
`LotArea:GarageArea`        -4.636e-03  1.289e-03  -3.596 0.000330 ***
`YearBuilt:MasVnrArea`       1.483e+00  2.021e-01   7.336 3.06e-13 ***
`YearBuilt:TotalBsmtSF`      6.293e-01  7.530e-02   8.357  < 2e-16 ***
`YearBuilt:Fireplaces`      -3.311e+02  5.074e+01  -6.526 8.35e-11 ***
`YearRemodAdd:FullBath`      3.917e+02  8.184e+01   4.786 1.81e-06 ***
`MasVnrArea:WoodDeckSf`      8.087e-02  3.110e-02   2.601 0.009369 **
`BsmtFinSF1:Fireplaces`      1.080e+01  3.036e+00   3.555 0.000386 ***
`TotalBsmtSF:Fireplaces`     1.087e+01  3.592e+00   3.027 0.002500 **
`TotalBsmtSF:WoodDeckSf`    -9.241e-03  1.327e-02  -0.696 0.486345
`FullBath:Bedroom`           9.994e+03  1.505e+03   6.639 3.96e-11 ***
`FullBath:Fireplaces`        1.484e+04  2.586e+03   5.740 1.08e-08 ***
`FullBath:GarageArea`        6.860e+01  8.083e+00   8.488  < 2e-16 ***
`HalfBath:Kitchen`          -2.683e+04  5.782e+03  -4.640 3.69e-06 ***
`HalfBath:Fireplaces`        1.259e+04  2.273e+03   5.541 3.36e-08 ***
`Bedroom:Fireplaces`        -9.338e+03  1.641e+03  -5.689 1.44e-08 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 31460 on 2230 degrees of freedom
Multiple R-squared:  0.8602,    Adjusted R-squared:  0.8575
F-statistic: 319.2 on 43 and 2230 DF,  p-value: < 2.2e-16
```

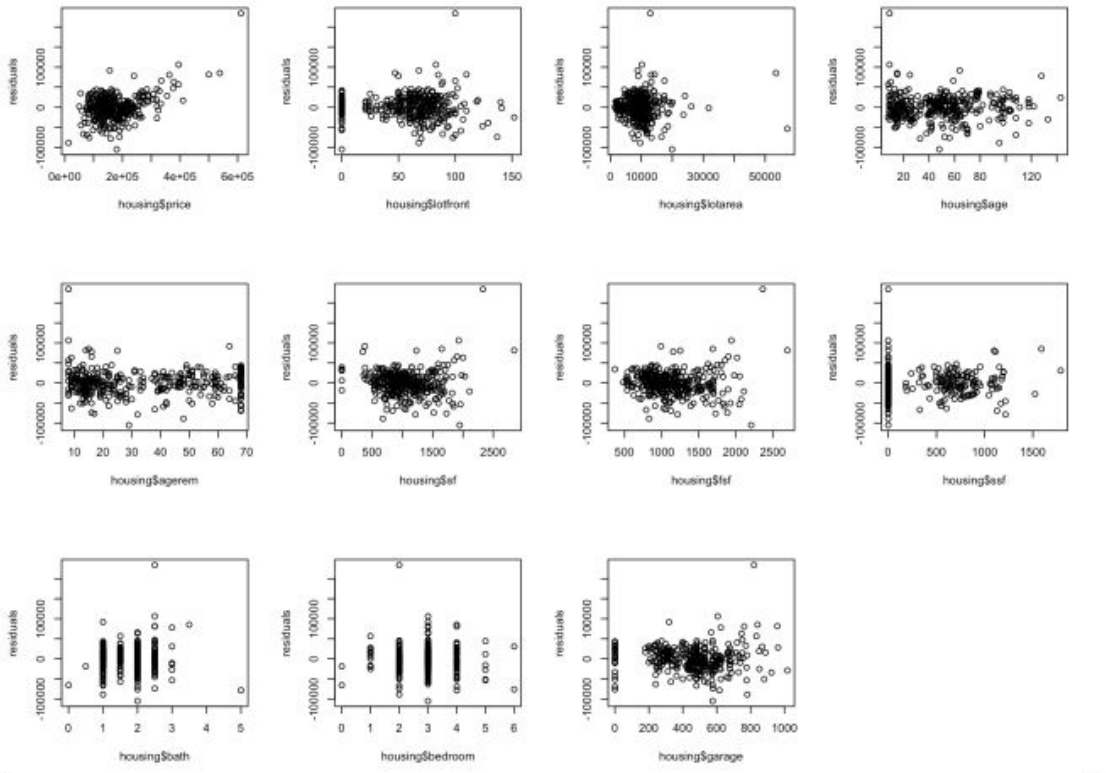
Assumptions

Residuals v. Y

No systematic pattern

Indicating constant

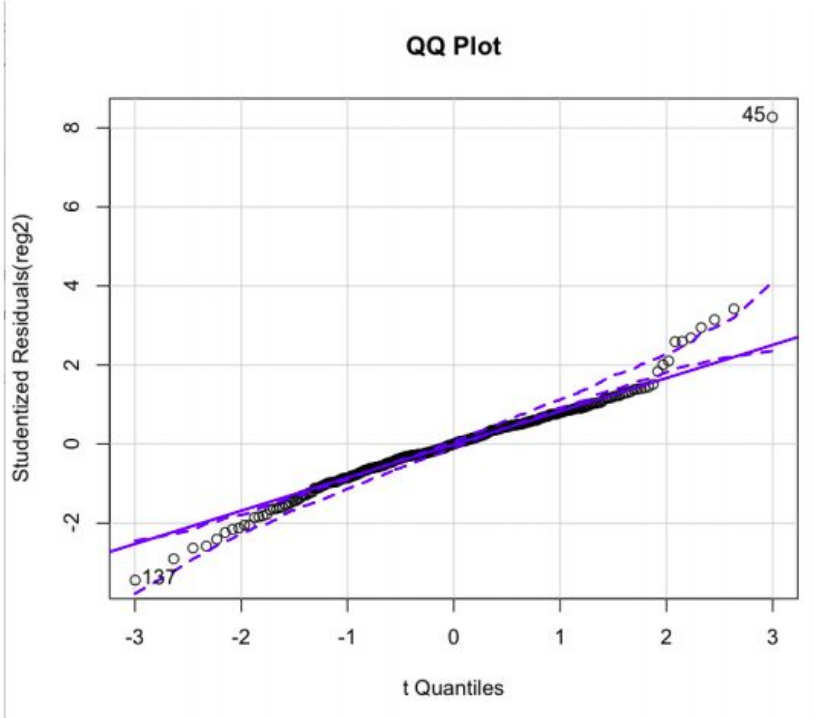
Variance and linearity



Assumptions(con’t)

QQ Plot: Showing normality

Straight, linear plot



Data Transformation

We didn’t preprocessing the data in the original first order linear model.

Centered the predictor variables when considered interaction terms.

The rounding error might be reduced.

Model Selection

Test whether certain coefficients is zero:

-F test for

Interpretations

Y-intercept: $16850

Positive Beta values will increase housing price: Lot Front, Square Footage,...

Negative Beta values will decrease housing price: Age of home, Last remodel

R-squared: 0.8602

Possible Problems

Most applicable to Ames, Iowa: Consider large economic differences such as

Crime rate, income, public transportation,....

Negative beta values in variables such as Bedrooms and Bathrooms

Removal of categorical variables in linear model