Chen-Fan (Annie) Lo
Ananya Mishra
Surya Narayana
Abhinav Pathak

Date:
19-Aug-2016

Analysis of US housing market rent

MSBA-6120

Introduction:

Background

- As a UK-based real estate company recently entering the US, it is important for us to understand the local market. A significant part of our revenue in the US is the rental income from the houses we build. So far, our leasing revenue is much worse than expected, and we struggle to increase our market share in the US. As a result, we performed an analysis to discover the major drivers that determine housing market rents in the US, gaining business insights that will help us increase revenue and expand our market share. As a real estate company with a diverse portfolio of properties, the analysis will help us design houses that are more likely to produce high rental income.

Problem Statement

- How can we help our sales and business managers to get good prediction on the rent of the properties and create a win-win situation in the market?

Objective

- Identify the potential market with the median rents in different regions.
- Identify the factors that determine and best predict housing rents in the US. We will incorporate these factors in construction projects, predict the rents of our properties, and design houses with desirable features.

- We are also interested in predicting rent levels to lease our houses at reasonable prices for customers.

Data Description

- **Time Period:** 2000-2012
- **Data Size:** The analysis is based on 7886 housing units in the US.
- **Summary of data**: Monthly rent levels varies between USD 438 and USD 3100. More summary of data can be found in Appendix (**Note 1**).
- **Factors considered:**
  1. **Region of houses**
     - Looking forward to expanding our business in the United States, we want to determine which region in the US we should target at. Regions are divided into four groups: Midwest, Northeast, South, and West.
  2. **Number of bedrooms in units**
     - Number of bedrooms indicates the number of people live in the units. We wanted to know if the number of bedroom is related to housing rents.
  3. **Number of units in a building**
     - Units is a matter of preference. Some prefer living in single houses and some like living in apartments. So, we considered this variable to check its influence on housing rents. This factor will help us decide what type of houses, for example, apartments or independent houses, we should build.
  4. **Median income of the area**
     - People tend to spend a large portion of their income on house rents. In order to predict the rent of our various housing units, we considered median income of the area as an important factor, as people with higher incomes can afford houses with higher rents, and vice versa.

5. **Age of houses**
    - New homes are typically more valuable than homes built many years ago. Thus, age of houses plays a major role in determining the rent of a house. This factor will help us find the depreciation rate of various housing units.
6. **Unemployment rate**
    - Areas with higher unemployment rates tend to have lower market rents. Therefore, we considered this parameter so that we can understand how unemployment rates affects house rents.
7. **Number of residents**
    - Houses that accommodate more residents usually have higher rents. Thus, we are interested in proving this view from the data.
8. **Tenure of stay**
    - Increase in rents per year is generally less when people continue to stay in the same house. We wanted to test if there is a correlation between tenure and house rent.

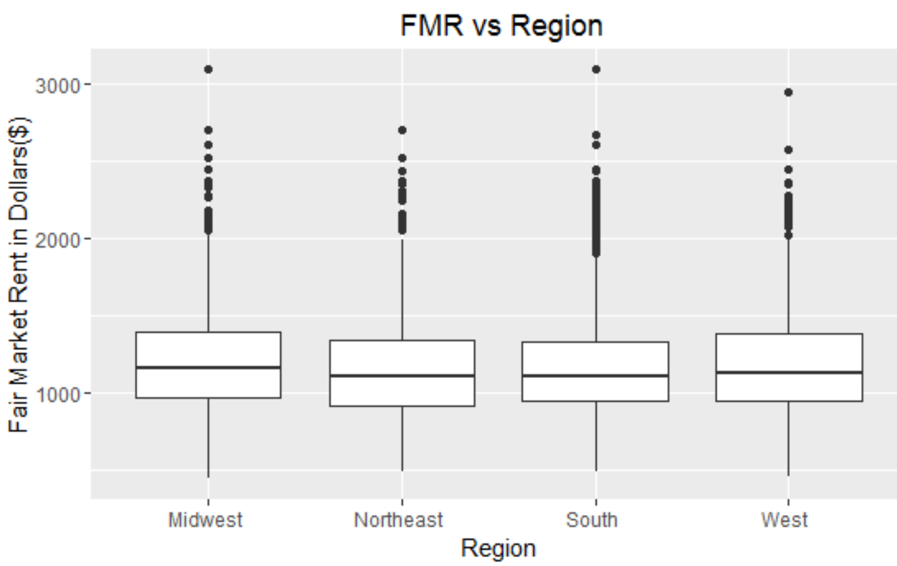
Analyses:

Which factors are related to rents?

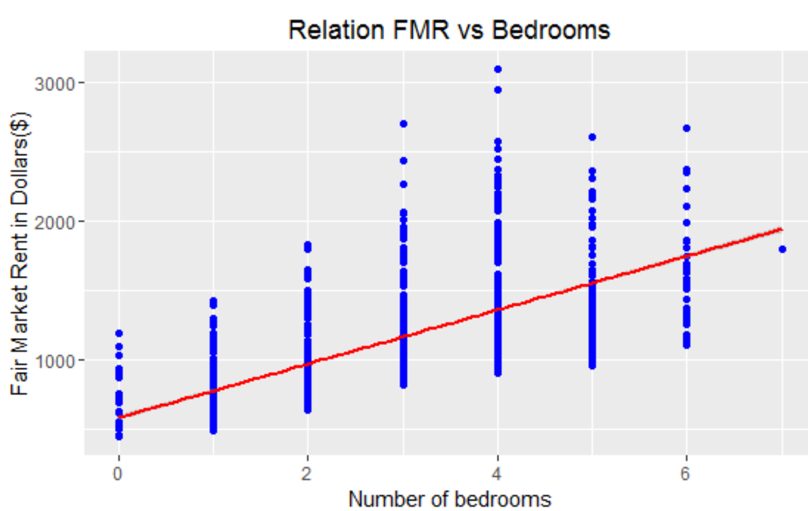
- First, we identified the correlation between individual factors and market rent levels in the US as follows:

1. **Rent levels in various regions**: Median and average rents are slightly higher in MidWest and West than in Northeast and South, as demonstrated in the below table and graph. All regions have rents above USD 2000. Most rents lie between USD 1000 to 1500 across all regions. Average rents is the highest in MidWest. Further detailed differences between average rents in four regions are attached in Appendix **(Note 2)**. Since there is no significant difference between regions, we cannot rely on this data to determine which region should be our main focus. More information has to be considered, for example, costs to construct a house.

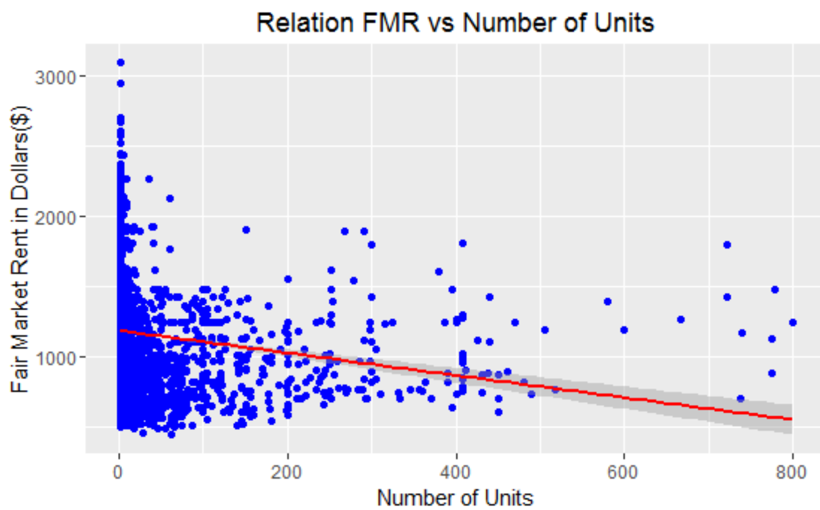
	Northeast	Midwest	South	West
Average rent	USD 1,154	USD 1,192	USD 1,165	USD 1,182
Median rent	USD 1,102	USD 1,153	USD 1,105	USD 1,125



2. As the **number of bedrooms** increases, rent also increases. Rents tend to be high when there are three, four, or five bedrooms in a house, and highest when there are four bedrooms in a house. Actually, 36.7% of the change in rents can be explained by the change in the number of bedrooms **(Note 3)**.



3. Houses with lesser **number of units** have most high and low rents at the same time. Around 1.7% of the change in rents is explained by the number of units **(Note 4)**. Thus, we should look into other factors that determine the rents when the number of units is small. Nevertheless, we discover that expensive rents mostly exist when the number of units is less than 50.



4. According to the data, **median income** is positively correlated to rents **(Note 5).** Median income explains around 24% change in the rents **(Note 6)**.
5. **Age of the house** does not have impact on rents. Thus, age of the house should not be our major concern **(Note 7)**. **Unemployment rate** also does not have strong relationship with rents. Houses in regions with high employment rates can also have high market rents **(Note 8)**.
6. There is positive relationship between the **number of residents** and rents. Most expensive rents lie in houses where have less than 7 people. The number of residents explain 11.21% change in rents **(Note 9)**. There is relationship between **Tenure** of 1, 2, and 3 years. Tenure explains 6% change in rents **(Note 10)**.

- We confirmed that the following six factors have impact on rents:
  - Region
  - Median income of the area
  - Number of bedroom in units

- Number of units in buildings
- Number of residents in units
- Tenure of stay

Optimized predictive model

Final four factors chosen for the predictive model:

- After we confirmed that there is no correlation between these factors **(Note 11)**, (as it may cause noises to our model), we built an optimized predictive model from some of these factors that we found most capable of predicting rent levels. We chose the model based on the extent of change in rents that can be explained by the factors and smallest noise **(Note 12)**.

  The final predictors we chose for our optimized predictive model are:
  - Median income of the area
  - Number of bedroom in units
  - Number of residents in units
  - Number of units in buildings
- Please refer to (**Note 13)** for our model selection process.

Predictive model:

log(Fair Market Rents) = -0.33 + 0.89*log(Median Income) + 0.17*(No. Of bedrooms) + 0.2*(No. Of units in buildings) + 0.1*(No. Of residents)

- The model explains 65% of change in rents.
- This model will help us predict rent levels when we design house projects.
- For every 10% increase in median income in an area, rent of a house goes up by 8.9%.
- If the number of bedrooms increases by 1, rent of a house goes up by 19%.
- If the number of units increases by 1, rent of a house goes up by 0.02%.
- If the number of residents increases by 1, rent of a house goes up by 1%.
- We have also tested the assumption of the model **(Note 12)**.

Scenario test:

- Assume:

Median Income	No. of bedrooms	No. of residents	No. of units
USD 55,555	2	4	10

- There is 90% chance that the average rent meeting the assumption will lie within USD 890.58 and USD 898.85 **(Note 14)**.
- If we randomly select a house that meets the assumption we made, 90% chance that the rent will fall in USD 684.94 and USD 1179.57 **(Note 14)**.

Conclusion:

From our analysis, we conclude that region, median income of the area, number of bedroom in units, number of units in buildings, number of residents in units, and tenor of stay have impact on market rents in the US. Thus, to increase our leasing income, we should build houses by incorporating these factors. On the other hand, age of houses and the unemployment rate of regions do not have relationship with rent levels.

We can also predict rent levels by using four factors - median income of the area, number of bedroom in units, number of units in buildings, and number of residents in units - with our prediction model. The model enables us to estimate approximate rent levels of houses in the US and thereby provide reasonable rents to customers, creating a win-win situation.

In addition, we found that Midwest has the highest median and average rents. We may consider to initiate our next project in the area. Meanwhile, more other factors need to be considered when determining the regions for expansion, such as labor and transportation costs of construction projects.

Limitations:

Decrease in the number of units and increase in the number of bedrooms are our limitations, as the rent for properties, such as bungalow or villa, can be impacted by other factors as follows:

- Swimming pools
- Gymnasium
- Other luxury features.

Appendix:

Statistics summary of data

<Note 1>

```
> summary(housedata)
   CONTROL              METRO3               LMED             FMR            BEDRMS     
 Length:7886        Length:7886        Min.   : 38500   Min.   : 438   Min.   :0.000  
 Class :character   Class :character   1st Qu.: 59541   1st Qu.: 949   1st Qu.:2.000  
 Mode  :character   Mode  :character   Median : 62258   Median :1117   Median :3.000  
                                       Mean   : 64673   Mean   :1172   Mean   :3.051  
                                       3rd Qu.: 70536   3rd Qu.:1349   3rd Qu.:4.000  
                                       Max.   :115300   Max.   :3100   Max.   :7.000  
     BUILT           TYPE        TENURE        NUNITS            PER           INFLATION    
 Min.   :2000   Min.   :1.000   1   :5235   Min.   :  1.00   Min.   : 1.000   Min.   :1.000  
 1st Qu.:2002   1st Qu.:1.000   2   :2096   1st Qu.:  1.00   1st Qu.: 2.000   1st Qu.:1.107  
 Median :2005   Median :1.000   3   :  59   Median :  1.00   Median : 2.000   Median :1.174  
 Mean   :2005   Mean   :1.085   NA's: 496   Mean   : 14.01   Mean   : 2.757   Mean   :1.189  
 3rd Qu.:2007   3rd Qu.:1.000               3rd Qu.:  1.00   3rd Qu.: 4.000   3rd Qu.:1.276  
 Max.   :2012   Max.   :9.000               Max.   :799.00   Max.   :11.000   Max.   :1.332  
                                                             NA's   :496                    
      AGE            REGION            Unemp_rate     
 Min.   : 4.00   Length:7886        Min.   :0.04000  
 1st Qu.: 9.00   Class :character   1st Qu.:0.04600  
 Median :11.00   Mode  :character   Median :0.05300  
 Mean   :11.31                      Mean   :0.05472  
 3rd Qu.:14.00                      3rd Qu.:0.05700  
 Max.   :16.00                      Max.   :0.09800  
```

Testing the relationship between Individual predictors and market rent levels

<Note 2>

```
> TukeyHSD(fit_region, conf.level = 0.9)
  Tukey multiple comparisons of means
    90% family-wise confidence level

Fit: aov(formula = FMR ~ REGION)

$REGION
                        diff        lwr       upr     p adj
Northeast -Midwest  -36.774341 -66.963241 -6.585442 0.0269857
South -Midwest      -25.518026 -49.174977 -1.861074 0.0644234
West -Midwest        -8.250845 -35.877811 19.376121 0.9030717
South -Northeast     11.256316 -15.547477 38.060108 0.7708270
West -Northeast      28.523496  -1.841636 58.888628 0.1367889
West -South          17.267181  -6.614254 41.148615 0.3467793
```

- From low p-values, we notice that there is a difference between mean rents of Midwest and Northeast region. Differences also tell us Midwest has the highest average rent.

<Note 3>

```
> summary(reg_bdrms)

Call:
lm(formula = FMR ~ BEDRMS)

Residuals:
    Min      1Q  Median      3Q     Max 
-643.67 -189.48  -63.44  119.50 1742.44 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  577.328      9.331   61.87   <2e-16 ***
BEDRMS       195.057      2.885   67.61   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 275.2 on 7884 degrees of freedom
Multiple R-squared:  0.367,	Adjusted R-squared:  0.3669 
F-statistic:  4571 on 1 and 7884 DF,  p-value: < 2.2e-16
```

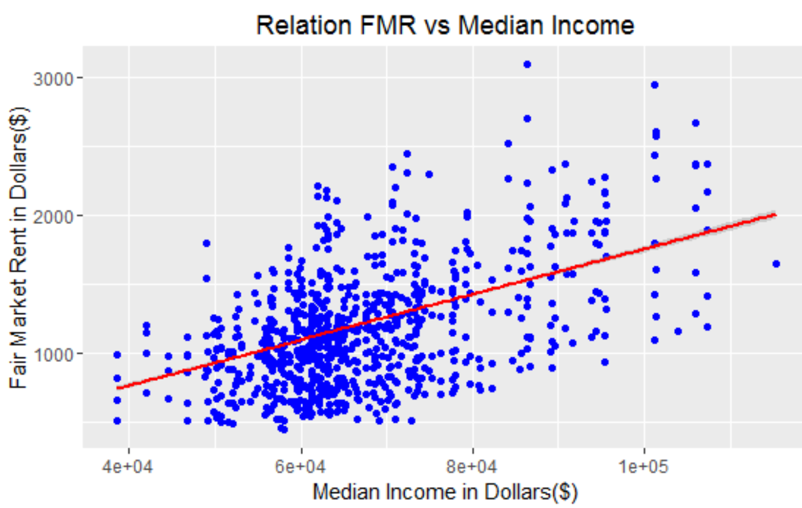
- From the summary of this linear model, we can infer that 36.7% of the variation in the Fair Market Rent is explained by the variability in Bedrooms, keeping other variables constant.
- We can say that for every additional bedroom the Fair Market Rent increases by $195.057.
- The p_values indicate the significance of relationship.

<Note 4>

```
> summary(reg_nunits)

Call:
lm(formula = FMR ~ NUNITS)

Residuals:
    Min      1Q  Median      3Q     Max 
-711.22 -231.46  -57.67  180.26 1917.26 

Coefficients:
              Estimate Std. Error t value Pr(>|t|)    
(Intercept) 1183.54070    3.98041   297.3   <2e-16 ***
NUNITS        -0.79731    0.06871   -11.6   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 343 on 7884 degrees of freedom
Multiple R-squared:  0.01679,	Adjusted R-squared:  0.01667 
F-statistic: 134.7 on 1 and 7884 DF,  p-value: < 2.2e-16
```

- Around 1.7% of the variance in Fair market rent is explained by the variability in number of units, keeping other variables constant.
- The p_values indicate the significance of relationship.

<Note 5>

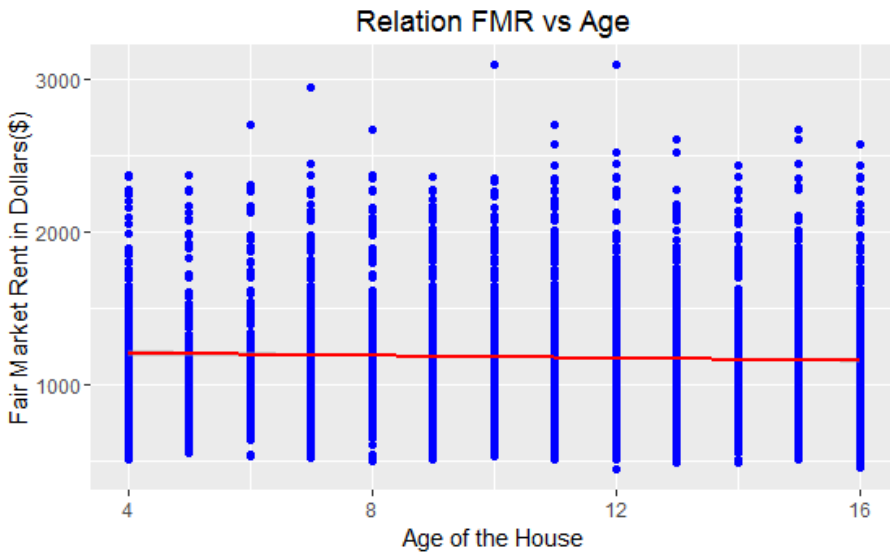


<Note 6>

```
> summary(reg_lmed)

Call:
lm(formula = FMR ~ LMED)

Residuals:
    Min      1Q  Median      3Q     Max 
-805.26 -172.46  -22.06  152.25 1570.39 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept) 1.041e+02  2.167e+01   4.801 1.61e-06 ***
LMED        1.652e-02  3.310e-04  49.904  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 301.5 on 7884 degrees of freedom
Multiple R-squared:  0.2401,	Adjusted R-squared:   0.24 
F-statistic:  2490 on 1 and 7884 DF,  p-value: < 2.2e-16
```

- Median income explains around 24% variance in the Fair market rent, keeping other variables constant.
- The p_values indicate the significance of relationship.

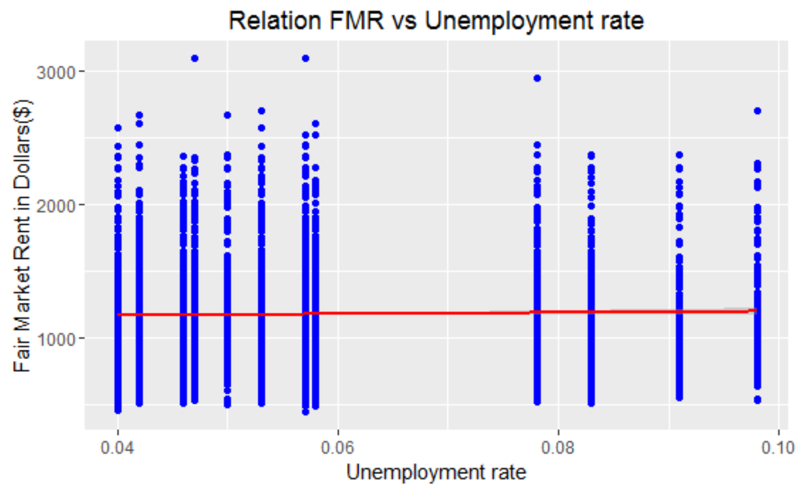
<Note 7>



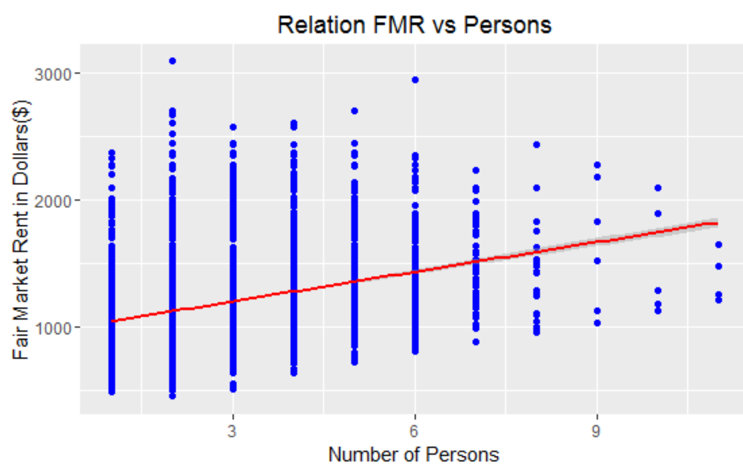

```
> reg_age <- lm(FMR~AGE)
> summary(reg_age)

Call:
lm(formula = FMR ~ AGE)

Residuals:
    Min      1Q  Median      3Q     Max 
-731.51 -225.94  -52.86  178.93 1930.49 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 1219.249     14.021   86.96  < 2e-16 ***
AGE           -4.145      1.191   -3.48 0.000503 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 345.6 on 7884 degrees of freedom
Multiple R-squared:  0.001534,	Adjusted R-squared:  0.001407 
F-statistic: 12.11 on 1 and 7884 DF,  p-value: 0.0005033
```

- Here we observe that the Age of the House just explains 0.1% of the variance in the Fair Market Rent, keeping other variables constant. This variable is not very helpful in regression and hence can be dropped from the model.

<Note 8>



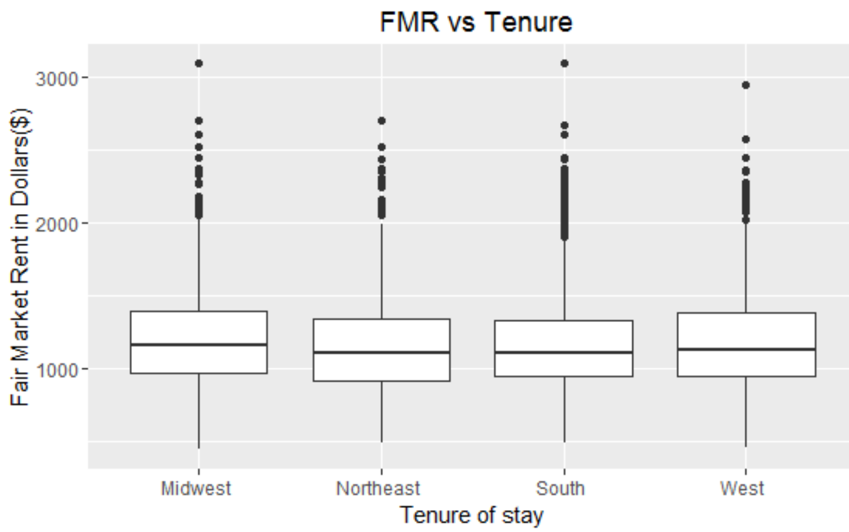
<Note 9>




```
> reg_per <- lm(FMR~PER)
> summary(reg_per)

Call:
lm(formula = FMR ~ PER)

Residuals:
    Min      1Q  Median      3Q     Max 
-697.47 -232.39  -51.58  165.69 1979.61 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept)  964.229      7.997  120.58   <2e-16 ***
PER           78.079      2.557   30.54   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 324.7 on 7388 degrees of freedom
  (496 observations deleted due to missingness)
Multiple R-squared:  0.1121,	Adjusted R-squared:  0.112 
F-statistic: 932.7 on 1 and 7388 DF,  p-value: < 2.2e-16
```

- Number of persons explain 11.21% variability in Fair market rent, keeping other variables constant.
- The p-values indicate the significance of relationship.

<Note 10>

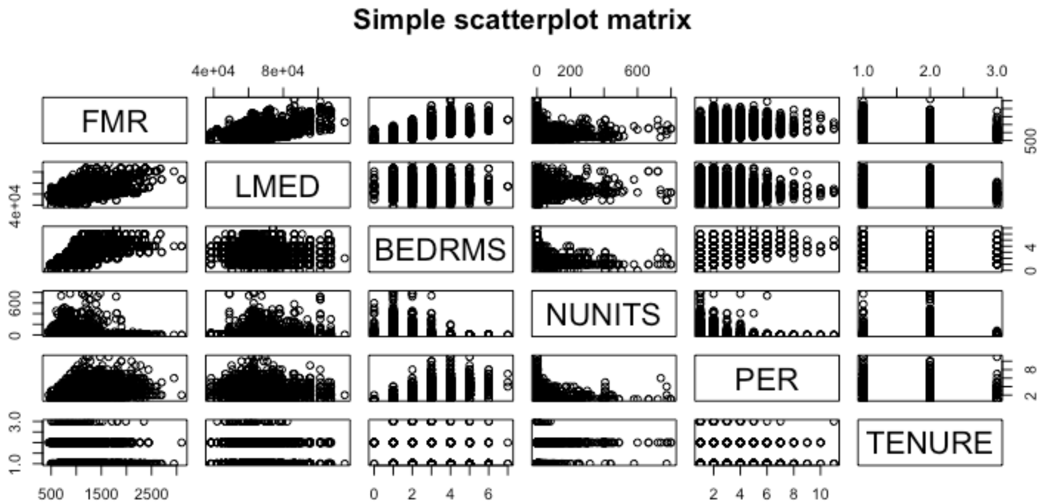



```
> summary(a)

Call:
lm(formula = FMR ~ factor(TENURE))

Residuals:
    Min      1Q  Median      3Q     Max 
-742.27 -240.01  -48.01  196.73 2051.99 

Coefficients:
                Estimate Std. Error t value Pr(>|t|)    
(Intercept)      1233.27       4.62 266.947  < 2e-16 ***
factor(TENURE)2  -185.26       8.64 -21.442  < 2e-16 ***
factor(TENURE)3  -156.27      43.76  -3.571 0.000358 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 334.3 on 7387 degrees of freedom
  (496 observations deleted due to missingness)
Multiple R-squared:  0.05926,	Adjusted R-squared:  0.05901 
F-statistic: 232.7 on 2 and 7387 DF,  p-value: < 2.2e-16
```

- Tenure explains around 6% variance in Fair market rents, keeping other variables constant.
- The p-values indicate the significance of relationship.

Collinearity Test

<Note 11>




```
> cor(housedata[sapply(housedata,is.numeric)])
                 LMED         FMR       BEDRMS       BUILT         TYPE       NUNITS PER
LMED        1.00000000  0.48995395 -0.021757813  0.01723203 -0.041091757  0.107571026  NA
FMR         0.48995395  1.00000000  0.605813787  0.03916775 -0.082145832 -0.129585731  NA
BEDRMS     -0.02175781  0.60581379  1.000000000  0.02560144 -0.073754831 -0.343449708  NA
BUILT       0.01723203  0.03916775  0.025601443  1.00000000 -0.061615900  0.069747669  NA
TYPE       -0.04109176 -0.08214583 -0.073754831 -0.06161590  1.000000000 -0.003181178  NA
NUNITS      0.10757103 -0.12958573 -0.343449708  0.06974767 -0.003181178  1.000000000  NA
PER                 NA          NA           NA          NA           NA           NA   1
INFLATION  -0.01164860 -0.03911193 -0.028305311 -0.99237316  0.064721449 -0.067356405  NA
AGE        -0.01723203 -0.03916775 -0.025601443 -1.00000000  0.061615900 -0.069747669  NA
Unemp_rate  0.03681702  0.02186802 -0.002223883  0.68525866 -0.026784272  0.061891637  NA
             INFLATION         AGE   Unemp_rate
LMED       -0.01164860 -0.01723203  0.036817022
FMR        -0.03911193 -0.03916775  0.021868019
BEDRMS     -0.02830531 -0.02560144 -0.002223883
BUILT      -0.99237316 -1.00000000  0.685258662
TYPE        0.06472145  0.06161590 -0.026784272
NUNITS     -0.06735640 -0.06974767  0.061891637
PER                 NA          NA           NA
INFLATION   1.00000000  0.99237316 -0.599044178
AGE         0.99237316  1.00000000 -0.685258662
Unemp_rate -0.59904418 -0.68525866  1.000000000
```

- From the matrix, we see that none of the correlation coefficients are greater than 0.7. The highest correlation is between PER(no. of persons) and BEDRMS(no. of bedrooms).

Model Selection

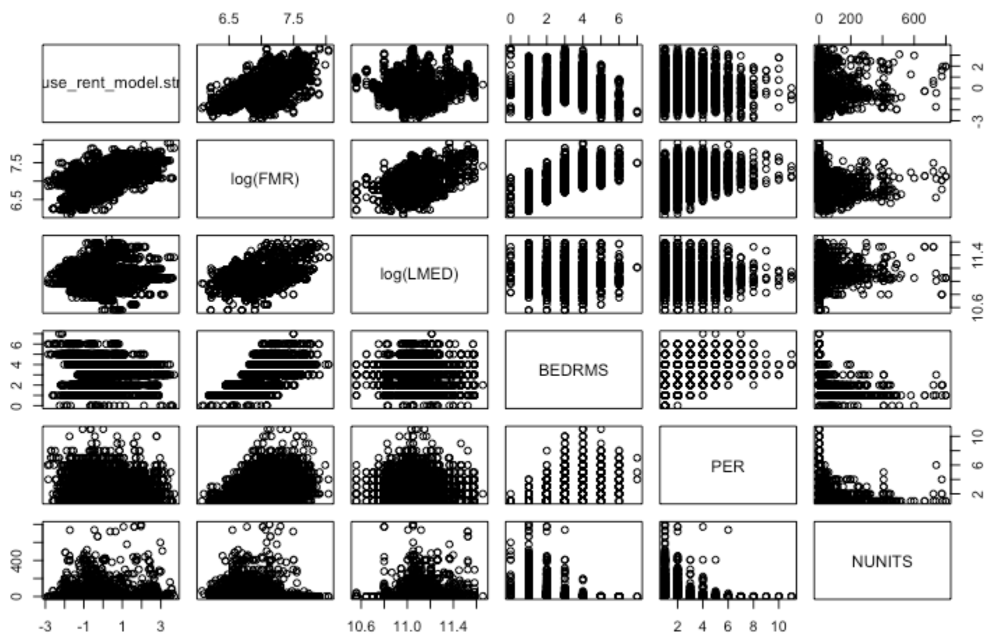
<Note 12> - the optimal model

```
> summary(house_rent_model2)

Call:
lm(formula = log(FMR) ~ log(LMED) + BEDRMS + NUNITS + PER)

Residuals:
     Min       1Q   Median       3Q      Max 
-0.48965 -0.11037 -0.02117  0.09715  0.62880 

Coefficients:
              Estimate Std. Error t value Pr(>|t|)    
(Intercept) -3.349e+00  1.403e-01 -23.876  < 2e-16 ***
log(LMED)    8.867e-01  1.268e-02  69.925  < 2e-16 ***
BEDRMS       1.746e-01  2.170e-03  80.467  < 2e-16 ***
NUNITS       2.011e-04  3.653e-05   5.505 3.80e-08 ***
PER          1.016e-02  1.528e-03   6.649 3.14e-11 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.1703 on 7881 degrees of freedom
Multiple R-squared:  0.6527,	Adjusted R-squared:  0.6525 
F-statistic:  3703 on 4 and 7881 DF,  p-value: < 2.2e-16
```

- We notice that there is no significant change in R-squared and standard errors in comparison to the Model 4.
- p-values for log(LMED)(for example): Probability of getting the t-value as high as 69.92, if there was no relationship between median income and fair market rent, keeping number of bedrooms, number of persons and no of units fixed.
- Estimates of log(LMED): For every 10% increase in median income in an area, the expected ratio of the two geometric means for free market rent will be 1.088 (1.1^0.89). OR, we expect about 8.8% increase in rent.
- Estimate of BEDRMS: If the number of bedrooms increase by 1, we expect to see about 19% (since exp(0.175) = 1.19) increase in rent.
- Estimate of NUNITS: If the number of units in a building increase by 1, we expect to see about 0.02% (since exp(0.0002) = 1.0002) increase in rent.
- Estimate of PER: If the number of persons in a house increase by 1, we expect to see about 1% (since exp(0.01) = 1.01) increase in rent.

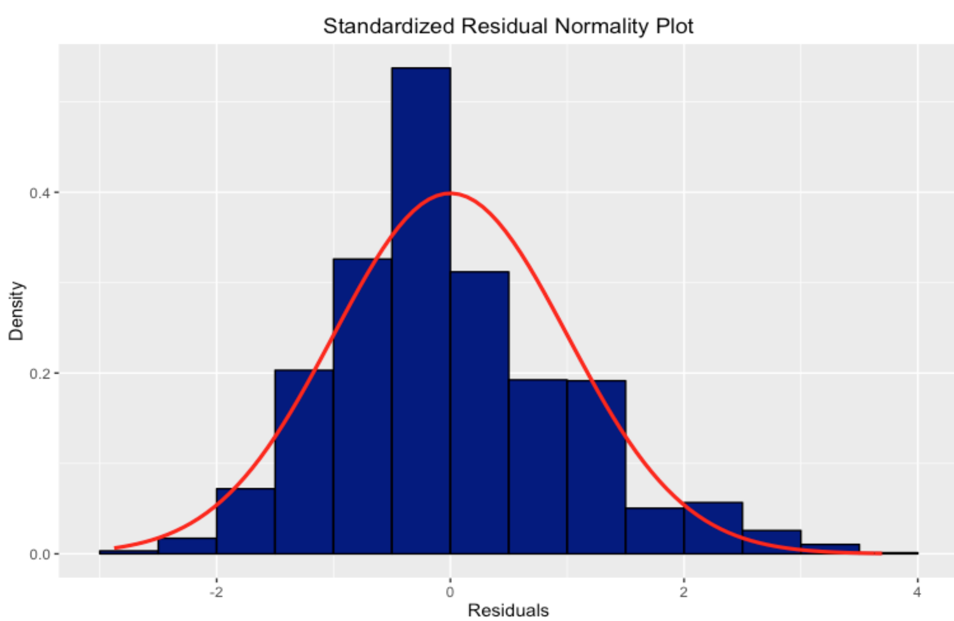
Assumption Test

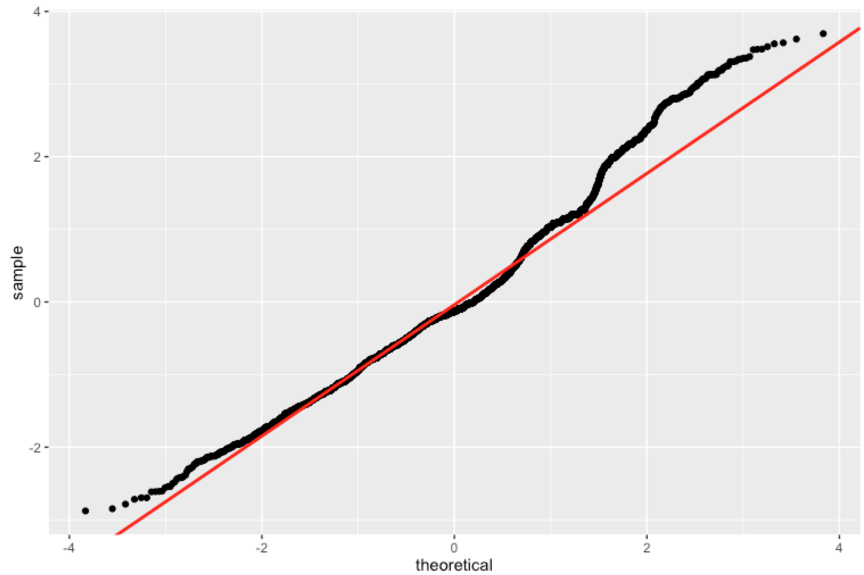
1. **Test of constant standard residuals & mean value of standard residuals**



- Our assumptions that standard residuals of the model are constant and have mean value of 0 are violated here. This needs to be checked or taken care of when this model is applied in real situations.

2. Normality test of standard residuals





- Our assumptions that standard residuals of the model are normally distributed are violated here. This needs to be checked or taken care of when this model is applied in real situations.

<Note 13> - Other models we have tested

Model 1:

```
Call:
lm(formula = FMR ~ LMED + BEDRMS + NUNITS + PER + TENURE + REGION)

Residuals:
    Min      1Q  Median      3Q     Max 
-578.14 -133.66  -47.25  101.51 1386.03 

Coefficients:
                  Estimate Std. Error t value Pr(>|t|)    
(Intercept)     -5.647e+02  2.017e+01 -28.002  < 2e-16 ***
LMED             1.684e-02  2.346e-04  71.771  < 2e-16 ***
BEDRMS           1.955e+02  2.907e+00  67.254  < 2e-16 ***
NUNITS           1.968e-01  4.568e-02   4.308 1.67e-05 ***
PER              1.200e+01  1.917e+00   6.262 4.00e-10 ***
TENURE           9.631e+00  5.654e+00   1.703   0.0885 .  
REGIONNortheast  9.212e-01  8.103e+00   0.114   0.9095    
REGIONSouth      2.438e+00  6.346e+00   0.384   0.7008    
REGIONWest       1.320e+01  7.408e+00   1.781   0.0749 .  
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 212.3 on 7877 degrees of freedom
Multiple R-squared:  0.6235,	Adjusted R-squared:  0.6232 
F-statistic:  1631 on 8 and 7877 DF,  p-value: < 2.2e-16
```

- From the above regressed model, Region and TENURE are found to be insignificant variables based on p-values. Model explains 62.3% of the change in Rent.

Model 2:

```
> summary(house_rent_model2)

Call:
lm(formula = FMR ~ LMED + BEDRMS + NUNITS + PER)

Residuals:
    Min      1Q  Median      3Q     Max 
-578.02 -136.08  -49.09  101.59 1391.07 

Coefficients:
              Estimate Std. Error t value Pr(>|t|)    
(Intercept) -5.443e+02  1.694e+01 -32.123  < 2e-16 ***
LMED         1.684e-02  2.345e-04  71.821  < 2e-16 ***
BEDRMS       1.936e+02  2.706e+00  71.558  < 2e-16 ***
NUNITS       2.027e-01  4.556e-02   4.448 8.77e-06 ***
PER          1.239e+01  1.905e+00   6.504 8.30e-11 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 212.4 on 7881 degrees of freedom
Multiple R-squared:  0.6232,	Adjusted R-squared:  0.623 
F-statistic:  3259 on 4 and 7881 DF,  p-value: < 2.2e-16
```

- This model has all the low p values with no significant change in r square and standard error from the previous model(Model 1).

Model 3:

```
> summary(house_rent_model3)

Call:
lm(formula = FMR ~ LMED + BEDRMS + PER)

Residuals:
    Min      1Q  Median      3Q     Max 
-569.17 -134.86  -50.96  102.06 1389.28 

Coefficients:
              Estimate Std. Error t value Pr(>|t|)    
(Intercept) -5.375e+02  1.689e+01  -31.82  < 2e-16 ***
LMED         1.696e-02  2.335e-04   72.62  < 2e-16 ***
BEDRMS       1.901e+02  2.590e+00   73.40  < 2e-16 ***
PER          1.226e+01  1.907e+00    6.43 1.35e-10 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 212.6 on 7882 degrees of freedom
Multiple R-squared:  0.6223,	Adjusted R-squared:  0.6221 
F-statistic:  4328 on 3 and 7882 DF,  p-value: < 2.2e-16
```

- After removing NUNITS(no. of units) variable, there is not much change in R-squared and standard error. Also, p-values remain significant.

Model 4:

```
> summary(house_rent_model)

Call:
lm(formula = log(FMR) ~ log(LMED) + BEDRMS * PER + NUNITS)

Residuals:
     Min       1Q   Median       3Q      Max 
-0.46069 -0.10659 -0.02909  0.10028  0.72347 

Coefficients:
              Estimate Std. Error t value Pr(>|t|)    
(Intercept) -3.770e+00  1.378e-01 -27.350   <2e-16 ***
log(LMED)    9.064e-01  1.237e-02  73.296   <2e-16 ***
BEDRMS       2.382e-01  3.658e-03  65.116   <2e-16 ***
PER          9.961e-02  4.459e-03  22.339   <2e-16 ***
NUNITS       3.523e-04  3.623e-05   9.724   <2e-16 ***
BEDRMS:PER  -2.608e-02  1.226e-03 -21.276   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.1656 on 7880 degrees of freedom
Multiple R-squared:  0.6716,	Adjusted R-squared:  0.6714 
F-statistic:  3223 on 5 and 7880 DF,  p-value: < 2.2e-16
```

- By adding a log and interaction term, R-square has improved by 5%, but here we cannot interpret estimates of 'no of bedrooms' and 'no of persons'.

Scenario Test

<Note 14>

```
> newdata <- data.frame(LMED = 55555, BEDRMS = 2, PER = 4, NUNITS = 10)
> conf_intvl <- predict(house_rent_model, newdata, interval="confidence", level = .90)
> exp(conf_intvl)
       fit      lwr      upr
1 898.8484 890.5799 907.1936
> prediction <- predict(house_rent_model, newdata, interval="predict", level = .90)
> exp(prediction)
       fit      lwr      upr
1 898.8484 684.9352 1179.569
```

References

- **Data Source:**

  U.S. Department of Housing and Urban Development - American Housing Survey: Housing Affordability Data System.

  https://www.huduser.gov/portal/datasets/hads/hads.html