# HW7 Multiple Regression Yuefei Chen

Yuefei Chen

2024-04-12

## The first part: Rent house Data

### Model development

Running the following code, we build a multiple regression model based on rent house data. Its independent variables "area", "rooms", "bathroom", "parking spaces", "hoa", "property tax", "fire insurance". The dependent variable is "rent amount".

```
library(readr)
house_data <- read_csv("Dataset/Rent_House_random_200_multi_regression.csv")
```

```
## Rows: 200 Columns: 11
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (3): floor, animal, furniture
## dbl (8): area, rooms, bathroom, parking_spaces, hoa, rent_amount, property_t...
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```
house_data <- house_data[, c(1:4, 6:11)]
str(house_data)
```

```
## tibble [200 x 10] (S3: tbl_df/tbl/data.frame)
##  $ area          : num [1:200] 120 45 50 35 204 177 15 70 180 180 ...
##  $ rooms         : num [1:200] 3 1 2 1 4 3 1 2 3 4 ...
##  $ bathroom      : num [1:200] 4 1 1 1 4 3 1 2 3 4 ...
##  $ parking_spaces: num [1:200] 3 1 1 0 2 4 0 1 2 2 ...
##  $ animal        : chr [1:200] "acept" "not acept" "acept" "acept" ...
##  $ furniture     : chr [1:200] "not furnished" "furnished" "not furnished" "not furnished" ...
##  $ hoa           : num [1:200] 1350 3000 226 260 0 2700 0 1800 700 2600 ...
##  $ rent_amount   : num [1:200] 5600 5520 750 1400 3440 6900 1200 4200 2700 2000 ...
##  $ property_tax  : num [1:200] 560 0 0 0 100 509 0 250 175 584 ...
##  $ fire_insurance: num [1:200] 71 70 10 18 62 89 16 55 40 26 ...
```

```
reg_data <- house_data[,-c(5,6)]
fit <- lm(rent_amount~area+rooms+bathroom+parking_spaces+hoa+property_tax+fire_insurance, data=reg_data)
fit
```

```
##
## Call:
## lm(formula = rent_amount ~ area + rooms + bathroom + parking_spaces +
##     hoa + property_tax + fire_insurance, data = reg_data)
##
## Coefficients:
##    (Intercept)            area           rooms        bathroom  parking_spaces
##       54.01250        -2.49739       -67.89201        66.10065        33.40648
##            hoa    property_tax  fire_insurance
##        0.32035        -0.03111        73.32810
```

### Model Acceptance

In the summary of the model, we focus on R squared value, coefficients, and P-value of each coefficient. The R-squared value is 0.9835 and Adjust R-squared value is 0.9829. It shows there is a high proportion of variance in the dependent variable can be explained by the independent variables. The result of coefficient is shown in the following table. P-value result shows that the "area", "hoa", and "fire insurance" are variables which have a significant relationship with the "rent amount" variable. In addition, we use anova to compare full model and reduced model. The result shows that only keeping "area", "hoa", and "fire insurance" does not improve the model performance. Therefore, we use stepAIC to find an optimal model. It contains "area", "rooms", "bathroom", "hoa", "fire insurance".

```
summary(fit)
```

```
##
## Call:
## lm(formula = rent_amount ~ area + rooms + bathroom + parking_spaces +
##     hoa + property_tax + fire_insurance, data = reg_data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2170.3  -126.8    -9.6   109.1  4011.7
##
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)     54.01250   76.47149   0.706  0.48085
## area            -2.49739    0.63604  -3.926  0.00012 ***
## rooms          -67.89201   41.35463  -1.642  0.10229
## bathroom        66.10065   45.03771   1.468  0.14383
## parking_spaces  33.40648   35.91696   0.930  0.35349
## hoa              0.32035    0.03715   8.623 2.42e-15 ***
## property_tax    -0.03111    0.04246  -0.733  0.46466
## fire_insurance  73.32810    1.23463  59.393  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 453.1 on 192 degrees of freedom
## Multiple R-squared:  0.9835, Adjusted R-squared:  0.9829
## F-statistic:  1634 on 7 and 192 DF,  p-value: < 2.2e-16
```

```
coefficients(fit)
```

```
##    (Intercept)           area          rooms       bathroom parking_spaces
```

```
##     54.0124954     -2.4973875    -67.8920130     66.1006537     33.4064754
##            hoa   property_tax fire_insurance
##      0.3203464     -0.0311120     73.3281041
```

```r
library(MASS)
fit1 <- fit
fit2 <- lm(rent_amount ~ area + hoa + fire_insurance, data = reg_data)
# compare models
anova(fit1, fit2)
```

```
## Analysis of Variance Table
##
## Model 1: rent_amount ~ area + rooms + bathroom + parking_spaces + hoa +
##     property_tax + fire_insurance
## Model 2: rent_amount ~ area + hoa + fire_insurance
##   Res.Df      RSS Df Sum of Sq      F Pr(>F)
## 1    192 39425392
## 2    196 40564017 -4  -1138625 1.3863 0.2402
```

```r
step <- stepAIC(fit, direction="both")
```

```
## Start:  AIC=2454.32
## rent_amount ~ area + rooms + bathroom + parking_spaces + hoa +
##     property_tax + fire_insurance
##
##                   Df Sum of Sq       RSS    AIC
## - property_tax     1    110224  39535616 2452.9
## - parking_spaces   1    177638  39603031 2453.2
## <none>                          39425392 2454.3
## - bathroom         1    442317  39867709 2454.6
## - rooms            1    553432  39978824 2455.1
## - area             1   3165776  42591168 2467.8
## - hoa              1  15268698  54694090 2517.8
## - fire_insurance   1 724340651 763766043 3045.1
##
## Step:  AIC=2452.88
## rent_amount ~ area + rooms + bathroom + parking_spaces + hoa +
##     fire_insurance
##
##                   Df Sum of Sq       RSS    AIC
## - parking_spaces   1    145498  39681115 2451.6
## <none>                          39535616 2452.9
## - bathroom         1    432197  39967813 2453.1
## - rooms            1    508595  40044211 2453.4
## + property_tax     1    110224  39425392 2454.3
## - area             1   3974795  43510411 2470.0
## - hoa              1  16784818  56320434 2521.7
## - fire_insurance   1 787848804 827384420 3059.1
##
## Step:  AIC=2451.61
## rent_amount ~ area + rooms + bathroom + hoa + fire_insurance
##
##                   Df Sum of Sq       RSS    AIC
```

```
## <none>                        39681115 2451.6
## - rooms           1    524159  40205274 2452.2
## + parking_spaces  1    145498  39535616 2452.9
## - bathroom        1    657291  40338406 2452.9
## + property_tax    1     78084  39603031 2453.2
## - area            1   3903457  43584571 2468.4
## - hoa             1  17055295  56736410 2521.1
## - fire_insurance  1 788294198 827975313 3057.2
```

```
fit3 <- lm(rent_amount ~ area + rooms + bathroom + hoa + fire_insurance, data = reg_data)
```

### Residual Analysis

Two plots are used in these residual analysis. The first plot is QQ plot. We can conclude that most of residual points are located in a straight line. It satisfies normal distribution. Simlarily, the componet + residual plots tells that each variable satisfies the normal distribution. The regression can be regarded as normal distribution.

```
confint(fit3,level=0.95)
```

```
##                       2.5 %      97.5 %
## (Intercept)     -95.5543465 203.9133827
## area             -3.6077450  -1.3634035
## rooms          -146.6132114  15.2417940
## bathroom         -7.6949583 161.2625040
## hoa               0.2443655   0.3790055
## fire_insurance   71.2507593  75.9265335
```

```
fitted(fit3)
```

```
##          1          2          3          4          5          6          7
##  5511.5579  6039.6885   681.6406  1383.9164  4154.0107  7038.4675  1205.4123
##          8          9         10         11         12         13         14
##  4510.7949  2801.7961  2374.8555  1061.6334  5011.4386  2624.1868  1352.2636
##         15         16         17         18         19         20         21
##  1127.6547   825.6313  4895.2536  2455.6685  5272.1779  1097.7289   834.4511
##         22         23         24         25         26         27         28
##  4490.7649  6566.8034 10807.7098  3097.4758  1250.5806  1690.6988   922.5070
##         29         30         31         32         33         34         35
##  1290.5269  8693.0235  2028.7218  4643.2566  4360.6289  1458.3538  3767.4580
##         36         37         38         39         40         41         42
##  1868.4123 15301.4530   816.2507  1651.3093  2004.1224  1270.5657  7457.7811
##         43         44         45         46         47         48         49
##  3449.8608  3627.2732  3423.9529  3802.9174 14342.9920  1255.3540 17677.9807
##         50         51         52         53         54         55         56
##  1401.5203  1750.0622  2018.9243  1774.8381  5446.8039  2433.8874  1133.8417
##         57         58         59         60         61         62         63
##  1313.4446  1339.1832   935.8842  2091.9363   851.7560  3448.6622  6055.1788
##         64         65         66         67         68         69         70
##  2247.1027 10188.1025  1866.4088  2611.3949  1254.8880  1974.9853   702.3652
##         71         72         73         74         75         76         77
##   964.9327  3415.8088  2855.8657  1180.5566  4354.3874  2343.5899   925.4704
```

```
##         78         79         80         81         82         83         84
##  1066.2249  3162.9495  3674.5577  1306.5155   919.0199  1992.6905  1043.8500
##         85         86         87         88         89         90         91
##  1440.2634  1101.9778  7737.6622   763.8804  2698.0348  5687.3217  2969.2884
##         92         93         94         95         96         97         98
##   579.6789  9985.8610  2387.3611  2911.1075  2567.5887  1775.5807  1846.7575
##         99        100        101        102        103        104        105
##  1241.8012  4183.5312 15283.0616  2168.4036   972.0547  1523.8742  1995.8247
##        106        107        108        109        110        111        112
##  6132.6245  9645.3327  1338.2410  1895.7396  1540.6618  1628.6900  6868.6737
##        113        114        115        116        117        118        119
##  1618.3260  2687.0499  2586.4870  3540.1317  9661.3000  2247.1525  9881.9970
##        120        121        122        123        124        125        126
##  3503.6685   894.3027  7469.7123  2874.1025  1589.7760   856.4075   704.8975
##        127        128        129        130        131        132        133
##   831.9656  2807.0680  1982.5924   704.5241   947.3628  1174.0745  1623.8423
##        134        135        136        137        138        139        140
##  6039.6885 14846.0121  4502.7869  2119.3653   820.0898  3125.2719  6047.3660
##        141        142        143        144        145        146        147
##  3839.0976  1057.2564  3272.9863  1678.2828  4989.6206  3404.2812 10029.1990
##        148        149        150        151        152        153        154
##  1005.7626  2706.7486   954.8195  4150.9097  4090.8074  1252.9744  3402.8335
##        155        156        157        158        159        160        161
##   996.0957 12692.3978  2180.9557  8965.6967  6342.4183  4730.8165   961.6378
##        162        163        164        165        166        167        168
##  1621.1194  1300.6583  8962.5088  3335.7245  2553.2451  1989.7880  7105.0819
##        169        170        171        172        173        174        175
## 12734.5496  1656.6040 12011.6440  3130.5997  8334.2638  2991.7432  1045.8071
##        176        177        178        179        180        181        182
## 10870.3632  1286.8722  1143.5071  3619.3600  1148.3880  3486.3153  4027.6544
##        183        184        185        186        187        188        189
##  3834.0908  2179.0729  2833.8547  1364.0389 14411.8397  3354.7650  1883.0096
##        190        191        192        193        194        195        196
##  2152.9957  4378.3633  5482.1623 11572.6168  6635.9162  1248.6417   767.4640
##        197        198        199        200
## 12457.1312  1082.9657  3602.8115  1974.1104
```

```
residuals(fit3)
```

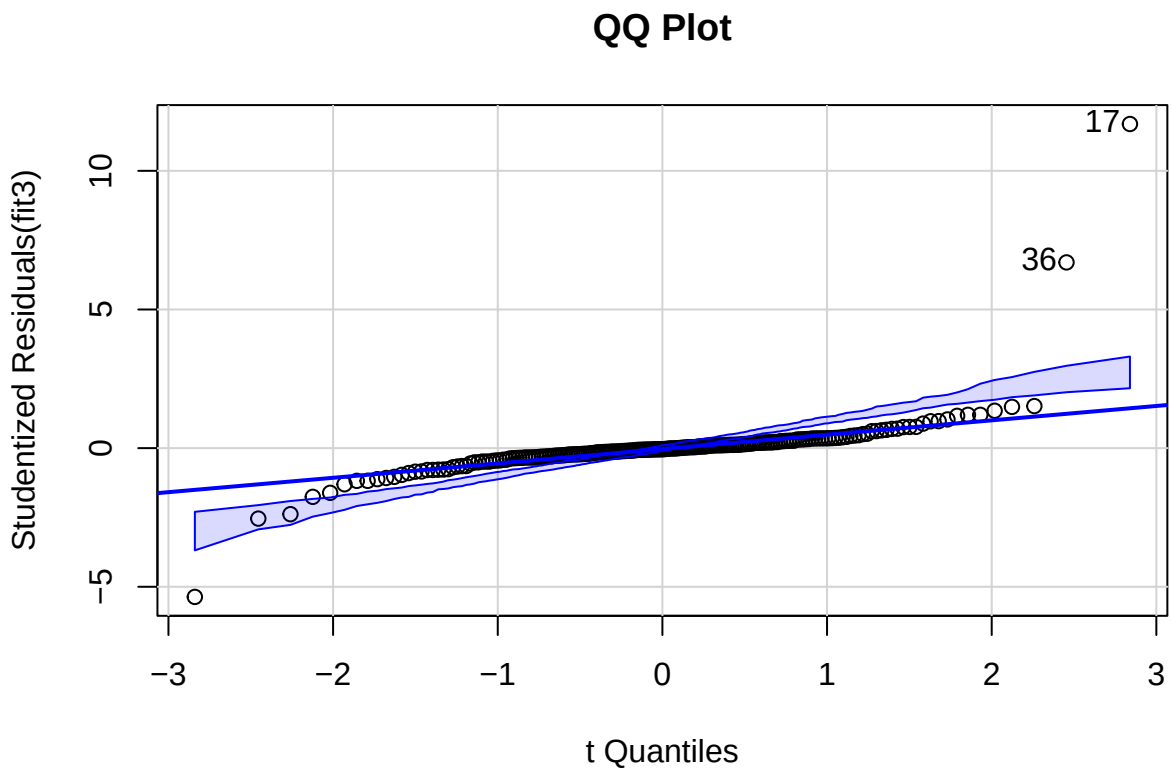
```
##           1           2           3           4           5           6
##    88.442092  -519.688515    68.359449    16.083648  -714.010710  -138.467475
##           7           8           9          10          11          12
##    -5.412311  -310.794918  -101.796062  -374.855542   -61.633434    46.561430
##          13          14          15          16          17          18
##  -124.186787  -152.263645   -61.654674   -17.631315  4004.746445  -147.668489
##          19          20          21          22          23          24
##   227.822100   -17.728868   165.548860  -290.764878   -66.803401  -307.709799
##          25          26          27          28          29          30
##  -297.475848  -230.580552  -200.698809    47.492983   -40.526926  -493.023510
##          31          32          33          34          35          36
##  -378.721778   156.743404   139.371087   141.646158   -67.457978  2631.587683
##          37          38          39          40          41          42
##  -301.453003   203.749273  -151.309333   195.877594  -120.565655  -157.781097
##          43          44          45          46          47          48
```

```
##    50.139241    27.726797    76.047118   447.082609   657.007996    -5.354039
##           49           50           51           52           53           54
##   322.019271  -101.520254    49.937764  -218.924255    85.161893    53.196104
##           55           56           57           58           59           60
##  -443.887365   -23.841698   -13.444561  -139.183153   -35.884170   -91.936271
##           61           62           63           64           65           66
##  -151.755952  -248.662227   274.821236  -207.102652 -2188.102529   233.591217
##           67           68           69           70           71           72
##    98.605078   -54.888007   -18.985313  -152.365221    85.067345   284.191240
##           73           74           75           76           77           78
##  -155.865700    19.443431   145.612614  -143.589880   -25.470391    13.775098
##           79           80           81           82           83           84
##    87.050471    25.442350    93.484491   -19.019913   437.309511   -43.849987
##           85           86           87           88           89           90
##   209.736567    38.022185   651.337792   -63.880433  -198.034819  -187.321719
##           91           92           93           94           95           96
##   530.711641   181.321088    14.138959     2.638869   -11.107505  -297.588740
##           97           98           99          100          101          102
##  -575.580672   -26.757526     8.198841 -1083.531184  -283.061593  -218.403571
##          103          104          105          106          107          108
##    27.945264   -23.874217     4.175282  -132.624486   154.667298   111.758978
##          109          110          111          112          113          114
##   -95.739641   159.338227   -28.689985  -221.673707   -58.326038   112.950055
##          115          116          117          118          119          120
##  -186.486990   -40.131736   338.700001    52.847540   118.003023    -3.668506
##          121          122          123          124          125          126
##    55.697333   270.287734   -74.102477    90.224017   -56.407505   -64.897482
##          127          128          129          130          131          132
##   -81.965566   342.932028   142.407647  -104.524063    52.637241    24.925525
##          133          134          135          136          137          138
##   176.157683  -519.688515   153.987947   147.213112   -19.365346  -100.089760
##          139          140          141          142          143          144
##    34.728062   -47.366011   160.902424  -107.256438   -22.986312   -28.282791
##          145          146          147          148          149          150
##    10.379416  -404.281222 -1029.199014    -5.762622    93.251355  -119.819482
##          151          152          153          154          155          156
##    -0.909698   -90.807429    47.025568   297.166490  -116.095663   307.602210
##          157          158          159          160          161          162
##   -80.955712  -165.696737   157.581704  -130.816472  -161.637786    93.880615
##          163          164          165          166          167          168
##   -20.658349   537.491164  -345.724512    46.754903   160.211988    94.918107
##          169          170          171          172          173          174
##   265.450425   -56.604012  -411.644041   -30.599708   165.736180    58.256838
##          175          176          177          178          179          180
##   -45.807148  -470.363188    43.127806   -93.507089  -219.360043   -48.388027
##          181          182          183          184          185          186
##    63.684654    72.345551    65.909170  -129.072871  -113.854663  -164.038895
##          187          188          189          190          191          192
##   588.160340  -354.764978   116.990406   -52.995710   121.636653   517.837746
##          193          194          195          196          197          198
##   427.383213  -785.916150   -98.641702    52.535969    42.868823   117.034259
##          199          200
##   397.188508    15.889605
```

```
library(car)
```

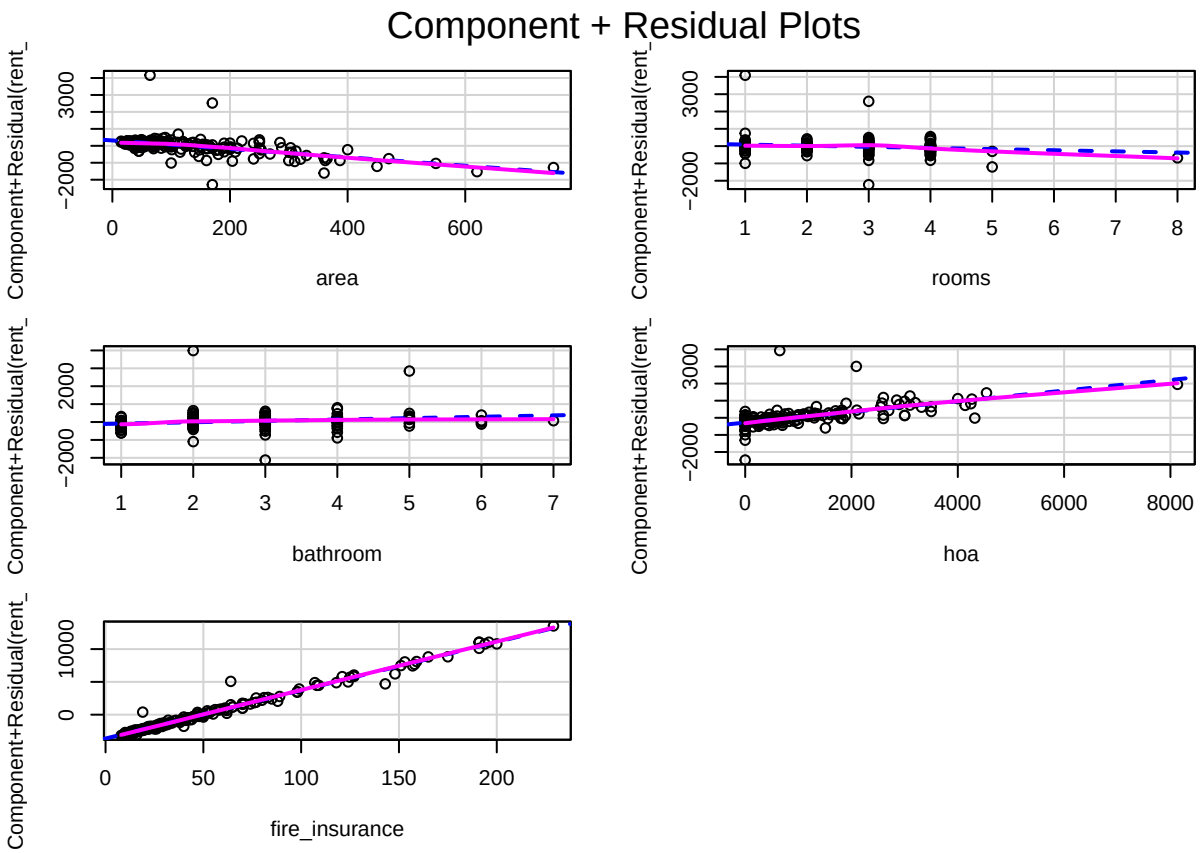
```
## Loading required package: carData
```

```
qqPlot(fit3, main="QQ Plot")
```




```
## [1] 17 36
```

```
crPlots(fit3)
```




### Prediction We set a data point with area = 120, rooms = 3, bathroom = 2, hoa = 0, fire insurance = 50, then the rent amount we predict is 3391.853.

```
predict.lm(fit3, data.frame(area = 120, rooms = 3, bathroom = 2, hoa = 0, fire_insurance = 50))
```

```
##        1
## 3391.853
```

### Model Accuracy

The accuracy is based on summary of the model and we also calculate the MSE and RMSE for the model. The R-squared value is 0.9834, and the Adjusted R-squared value is 0.983. The MSE is 198405.6 and RMSE is 445.4274.

```
summary(fit3)
```

```
##
## Call:
## lm(formula = rent_amount ~ area + rooms + bathroom + hoa + fire_insurance,
##     data = reg_data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -2188.1  -129.5   -17.7    94.1  4004.7
##
```

```
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)     54.17952   75.91965   0.714   0.4763
## area            -2.48557    0.56897  -4.369 2.04e-05 ***
## rooms          -65.68571   41.03272  -1.601   0.1110
## bathroom        76.78377   42.83330   1.793   0.0746 .
## hoa              0.31169    0.03413   9.131  < 2e-16 ***
## fire_insurance  73.58865    1.18538  62.080  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 452.3 on 194 degrees of freedom
## Multiple R-squared:  0.9834, Adjusted R-squared:  0.983
## F-statistic:  2296 on 5 and 194 DF,  p-value: < 2.2e-16
```

```
predictions <- predict(fit3, reg_data)
mse <- mean((reg_data$rent_amount - predictions)^2)
rmse <- sqrt(mse)
cat("MSE: ", mse, "\n")
```

```
## MSE:  198405.6
```

```
cat("RMSE: ", rmse, "\n")
```

```
## RMSE:  445.4274
```