Midterm Exam

RE6124019

2023-11-21

```
estate <- read.csv("Q5.csv")
```

可以看到共12列變量,涵蓋了房屋交易年份、地理位置、面積、價格等信息。

將Year設為因子變量。

```
estate$Year <- as.factor(estate$Year)
```

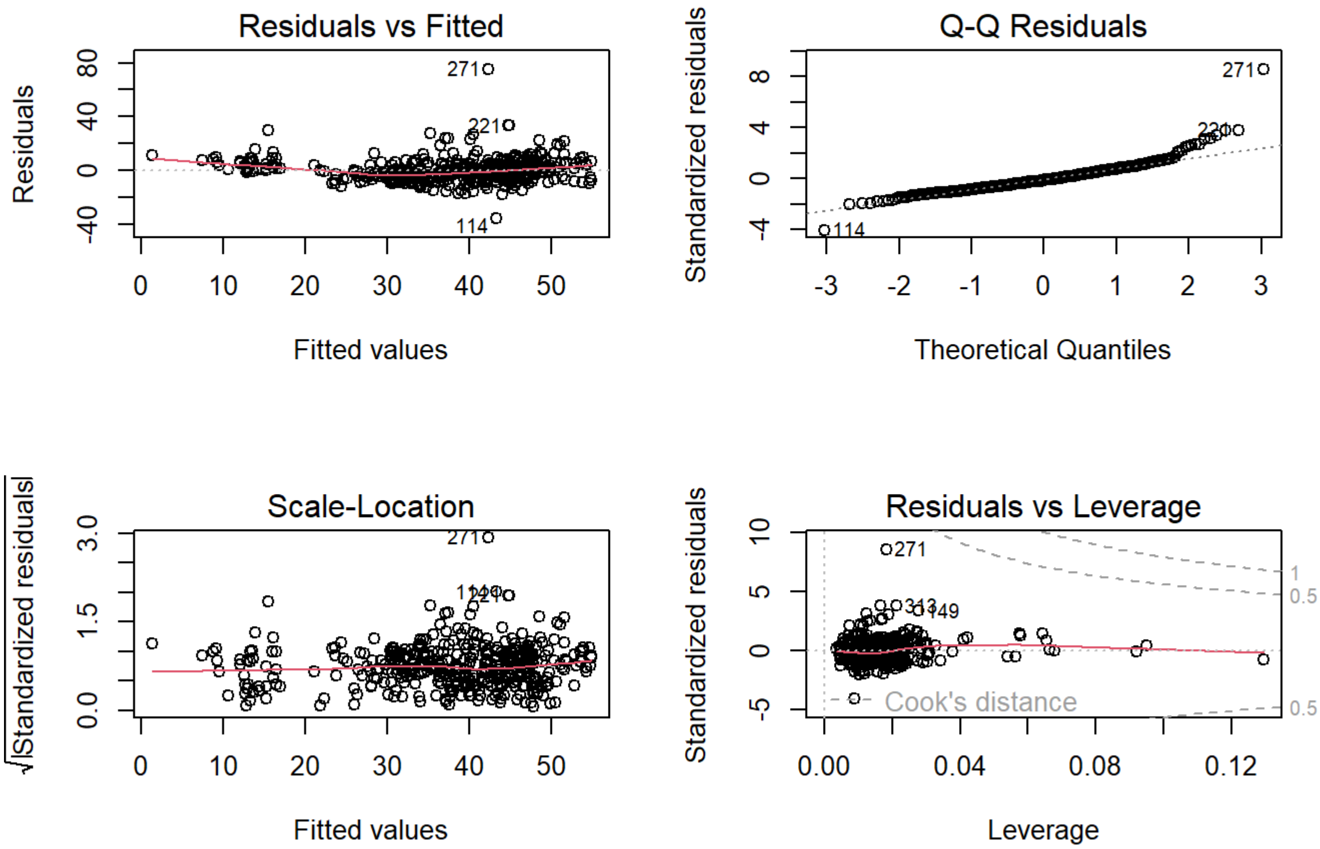
模型建立 應變數選擇房價Y,不進行變換。

構建包含所有數值預測變量的初始模型:

```
full_model <- lm(Y ~ . -Year, data = estate)
summary(full_model)
```

```
##
## Call:
## lm(formula = Y ~ . - Year, data = estate)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -35.664  -5.410  -0.966   4.217  75.193
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -1.444e+04  6.776e+03  -2.131  0.03371 *
## X1           5.146e+00  1.557e+00   3.305  0.00103 **
## X2          -2.697e-01  3.853e-02  -7.000 1.06e-11 ***
## X3          -4.488e-03  7.180e-04  -6.250 1.04e-09 ***
## X4           1.133e+00  1.882e-01   6.023 3.84e-09 ***
## X5           2.255e+02  4.457e+01   5.059 6.38e-07 ***
## X6          -1.242e+01  4.858e+01  -0.256  0.79829
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 8.858 on 407 degrees of freedom
## Multiple R-squared:  0.5824, Adjusted R-squared:  0.5762
## F-statistic: 94.59 on 6 and 407 DF,  p-value: < 2.2e-16
```

```
par(mfrow=c(2,2))
plot(full_model)
```



模型診斷 檢查模型殘差圖,未發現明顯違反假設的情況。

最終模型 逐步回歸選擇模型:

```
final_model <- step(full_model)
```

```
## Start:  AIC=1813.04
## Y ~ (X1 + X2 + X3 + X4 + X5 + X6 + Year) - Year
##
##         Df Sum of Sq   RSS    AIC
## - X6    1       5.1 31938 1811.1
## <none>              31933 1813.0
## - X1    1     857.0 32790 1822.0
## - X5    1    2008.2 33941 1836.3
## - X4    1    2846.0 34779 1846.4
## - X3    1    3064.5 34997 1849.0
## - X2    1    3843.9 35776 1858.1
##
## Step:  AIC=1811.11
## Y ~ X1 + X2 + X3 + X4 + X5
##
##         Df Sum of Sq   RSS    AIC
## <none>              31938 1811.1
## - X1    1     853.9 32792 1820.0
## - X5    1    2064.9 34003 1835.0
## - X4    1    2870.6 34808 1844.7
## - X2    1    3838.9 35777 1856.1
## - X3    1    6181.8 38119 1882.4
```

```
summary(final_model)
```

```
##
## Call:
## lm(formula = Y ~ X1 + X2 + X3 + X4 + X5, data = estate)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -35.623  -5.371  -1.020   4.244  75.346
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept) -1.596e+04  3.233e+03  -4.936 1.17e-06 ***
## X1           5.135e+00  1.555e+00   3.303  0.00104 **
## X2          -2.694e-01  3.847e-02  -7.003 1.04e-11 ***
## X3          -4.353e-03  4.899e-04  -8.887  < 2e-16 ***
## X4           1.136e+00  1.876e-01   6.056 3.17e-09 ***
## X5           2.269e+02  4.417e+01   5.136 4.36e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 8.848 on 408 degrees of freedom
## Multiple R-squared:  0.5823, Adjusted R-squared:  0.5772
## F-statistic: 113.8 on 5 and 408 DF,  p-value: < 2.2e-16
```

最終模型的自由度調整後R^2為0.5823。

結論 影響房價的主要預測變量為X1、X4、X5,且都呈現正相關。