Analysis

Dingyi Li

4/14/2021

Multiple Linear Regression

Data preview

Read in data

```
dt = read.csv("data&figures/dt.csv")
summary(dt)
```

```
##     State             County               HPI          Personal_Income
##  Length:2168        Length:2168        Min.   : 96.64   Min.   : 27415
##  Class :character   Class :character   1st Qu.:138.98   1st Qu.: 38812
##  Mode  :character   Mode  :character   Median :152.19   Median : 43976
##                                        Mean   :159.19   Mean   : 46355
##                                        3rd Qu.:175.13   3rd Qu.: 50669
##                                        Max.   :395.90   Max.   :229825
##  Poverty_Percentage   Population       HighSchoolLess   HighSchoolOnly
##  Min.   : 2.70      Min.   :   1129   Min.   : 1.50    Min.   : 7.80
##  1st Qu.: 9.70      1st Qu.:  18913   1st Qu.: 8.10    1st Qu.:29.10
##  Median :12.65      Median :  37068   Median :11.10    Median :33.90
##  Mean   :13.36      Mean   : 123587   Mean   :12.14    Mean   :33.64
##  3rd Qu.:16.02      3rd Qu.:  93783   3rd Qu.:15.12    3rd Qu.:38.60
##  Max.   :38.20      Max.   :5150233   Max.   :43.10    Max.   :54.50
##   SomeCollege     BachelorAndHigher Unemployment_Rate
##  Min.   :11.20   Min.   : 8.20     Min.   : 1.600
##  1st Qu.:27.90   1st Qu.:16.30     1st Qu.: 3.100
##  Median :31.00   Median :20.80     Median : 3.700
##  Mean   :31.08   Mean   :23.14     Mean   : 3.888
##  3rd Qu.:34.20   3rd Qu.:28.12     3rd Qu.: 4.500
##  Max.   :47.30   Max.   :75.30     Max.   :18.300
```

Correlation Check

```
cor(scale(as.matrix(dt[,c(7,8,9,10)])))
```

Education parameters

```
##                    HighSchoolLess HighSchoolOnly SomeCollege BachelorAndHigher
## HighSchoolLess          1.0000000      0.2768850  -0.3538971        -0.5983478
## HighSchoolOnly          0.2768850      1.0000000  -0.2113544        -0.7972454
## SomeCollege            -0.3538971     -0.2113544   1.0000000        -0.1358199
## BachelorAndHigher      -0.5983478     -0.7972454  -0.1358199         1.0000000
```

Histogram

```r
par(mfrow = c(3,3))
hist(dt$HPI)
hist(dt$Personal_Income)
hist(dt$Poverty_Percentage)
hist(dt$Population)
hist(dt$Unemployment_Rate)
hist(dt$BachelorAndHigher)
hist(dt$SomeCollege)
```
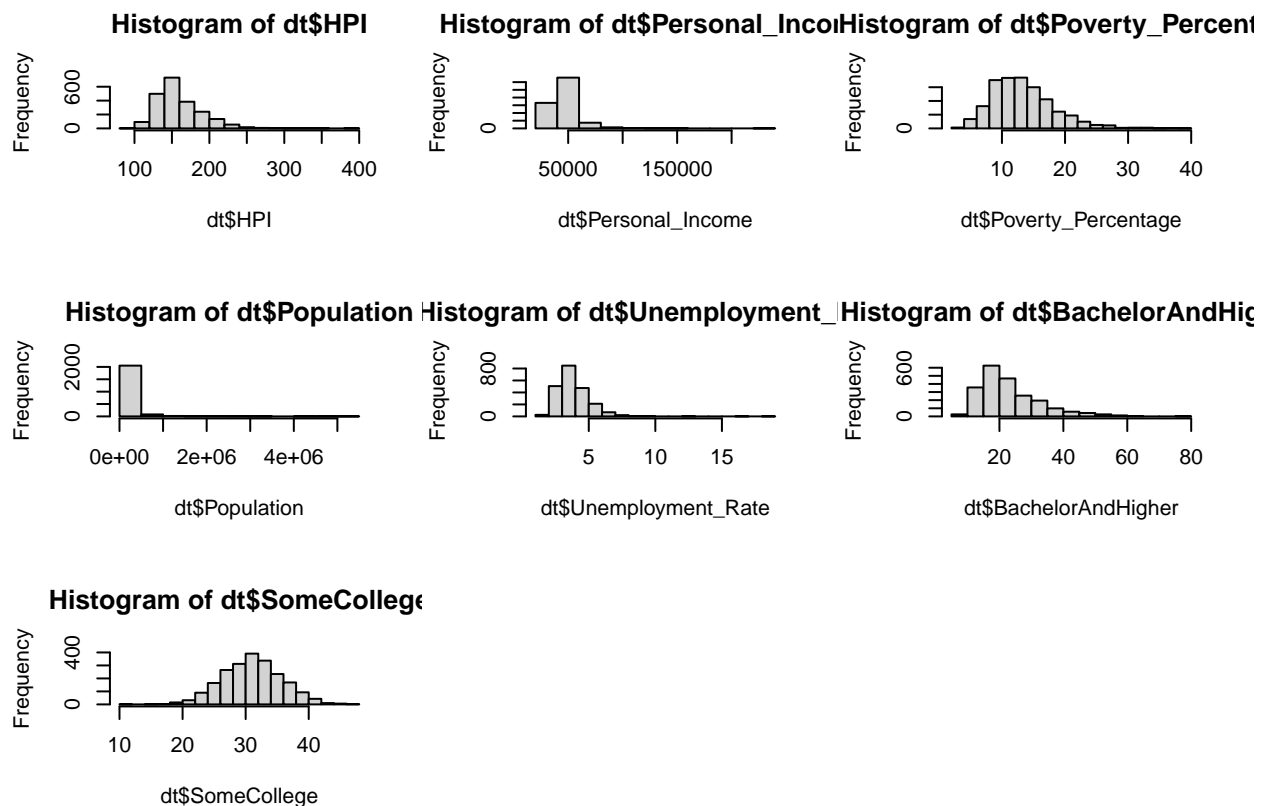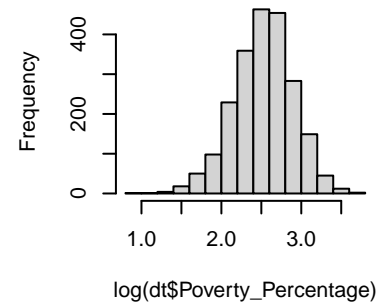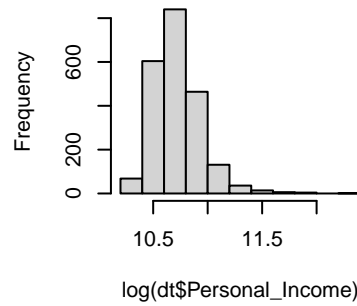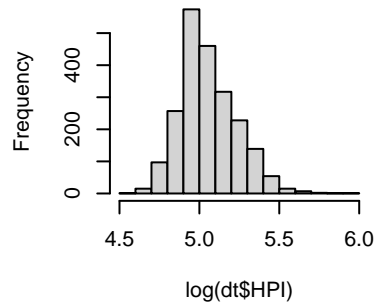

Histogram for logtransformation

```r
par(mfrow = c(2,3))
hist(log(dt$HPI))
hist(log(dt$Personal_Income))
hist(log(dt$Poverty_Percentage))
hist(log(dt$Population))
```

```
hist(log(dt$Unemployment_Rate))
hist(log(dt$BachelorAndHigher))
```

Histogram of log(dt$HPI)   Histogram of log(dt$Personal_Incstogram of log(dt$Poverty_Perce

Histogram of log(dt$Populatiostogram of log(dt$Unemploymentstogram of log(dt$BachelorAndH
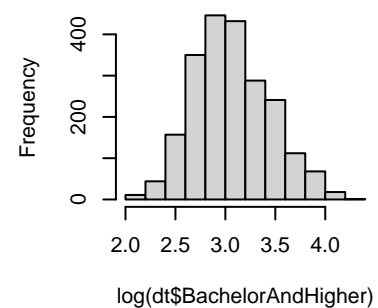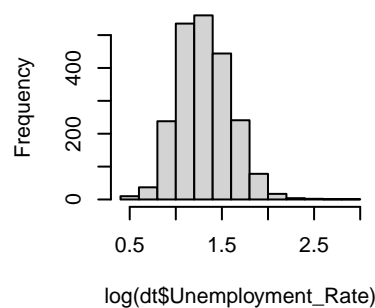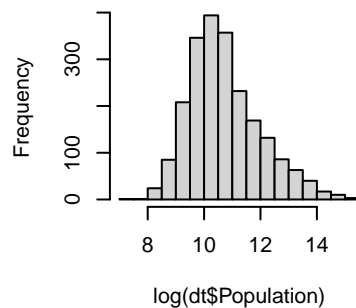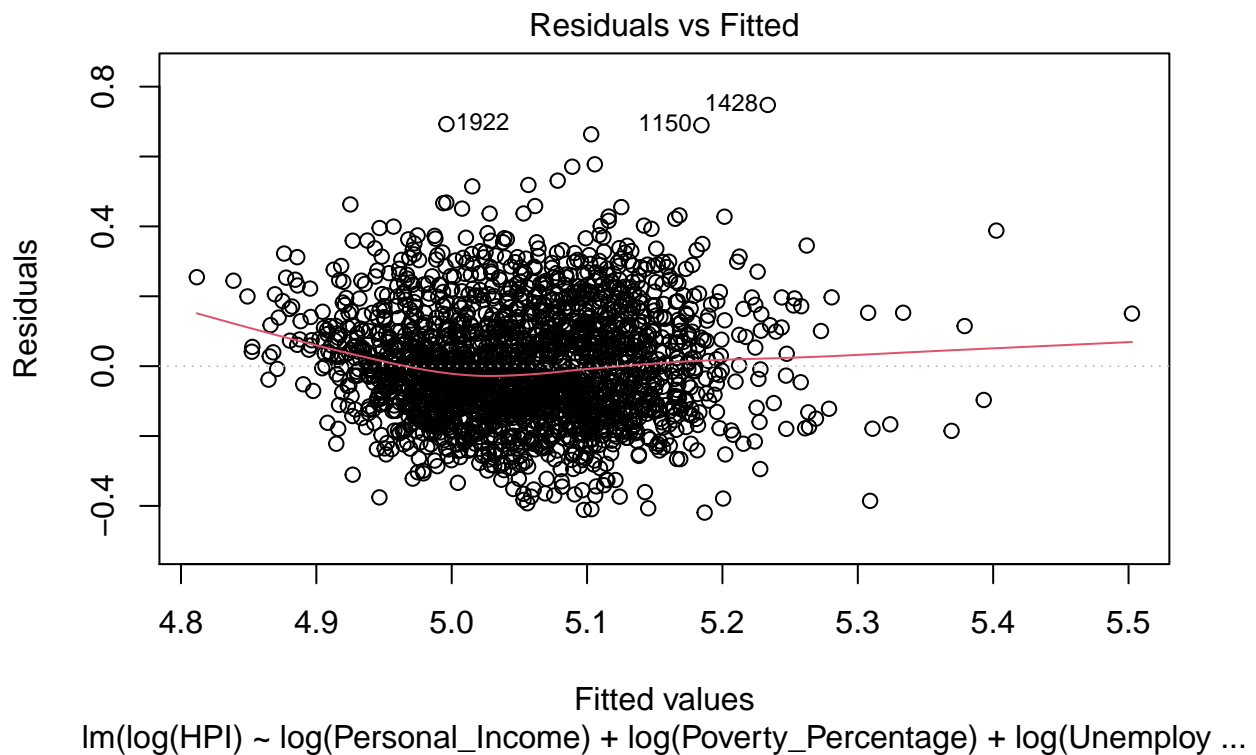
Model fitting

```
attach(dt)
m1 = lm(log(HPI)~log(Personal_Income)+log(Poverty_Percentage)+log(Unemployment_Rate)+log(Population)+Som
summary(m1)
```
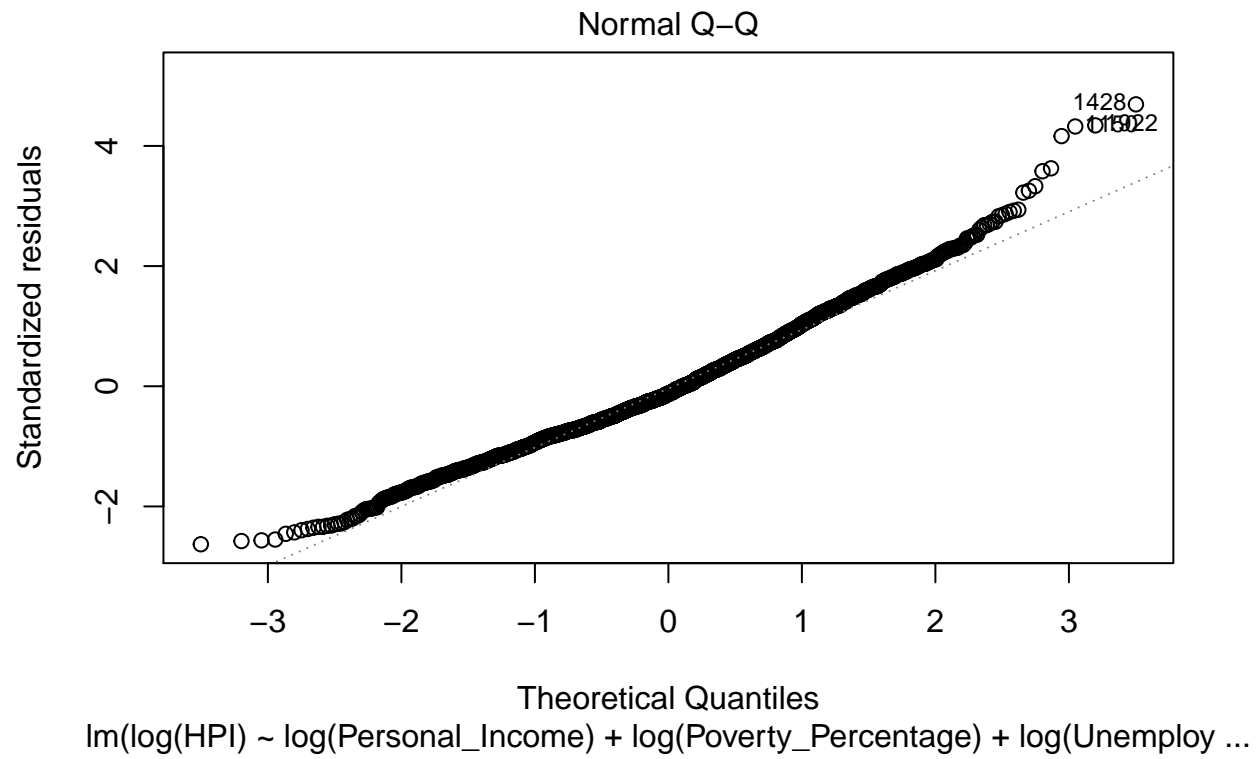
```
##
## Call:
## lm(formula = log(HPI) ~ log(Personal_Income) + log(Poverty_Percentage) +
##     log(Unemployment_Rate) + log(Population) + SomeCollege +
##     log(BachelorAndHigher))
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.41914 -0.11184 -0.02073  0.09933  0.74766
##
## Coefficients:
##                        Estimate Std. Error t value Pr(>|t|)
## (Intercept)           1.5982142  0.2873397   5.562 3.00e-08 ***
## log(Personal_Income)  0.2841528  0.0261091  10.883  < 2e-16 ***
```
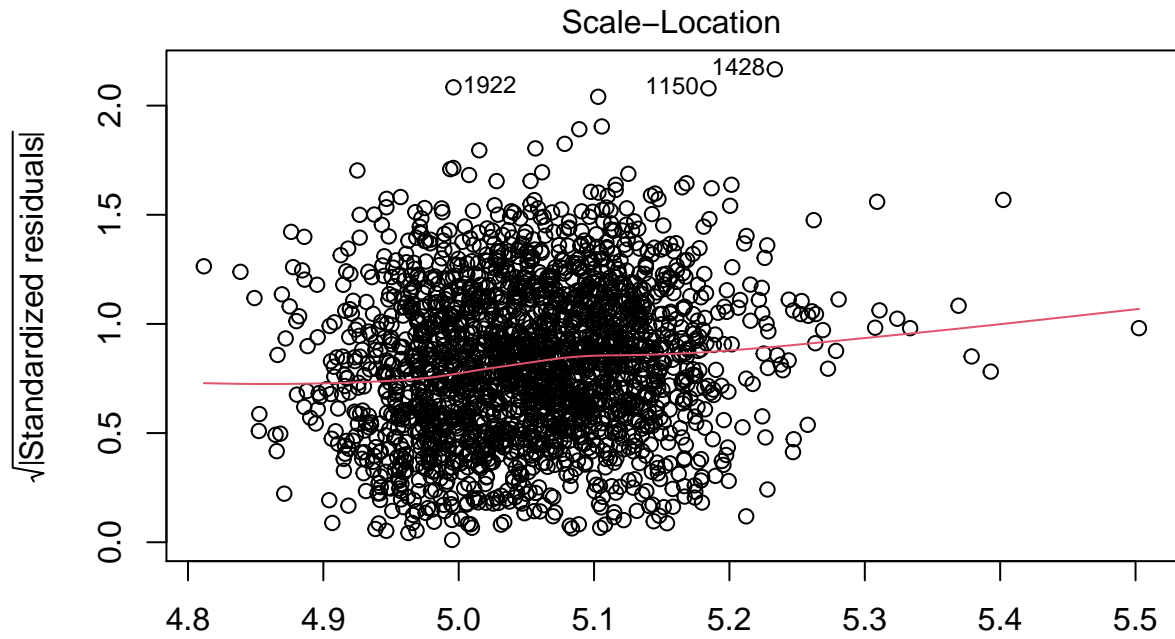
```
## log(Poverty_Percentage)   0.0658601  0.0140051    4.703 2.73e-06 ***
## log(Unemployment_Rate)   -0.0760410  0.0144298   -5.270 1.50e-07 ***
## log(Population)           0.0017443  0.0032963    0.529 0.596751
## SomeCollege               0.0052240  0.0007657    6.823 1.15e-11 ***
## log(BachelorAndHigher)    0.0530982  0.0145109    3.659 0.000259 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1599 on 2161 degrees of freedom
## Multiple R-squared:  0.1883, Adjusted R-squared:  0.186
## F-statistic: 83.53 on 6 and 2161 DF,  p-value: < 2.2e-16
```
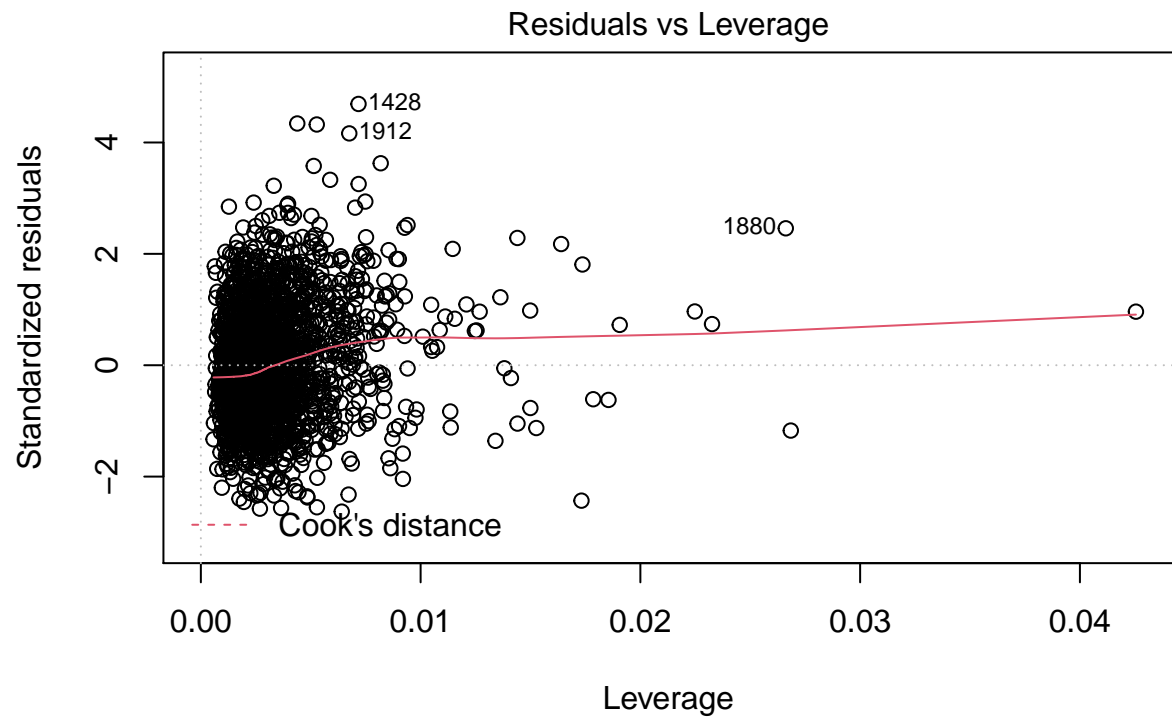
Diagnostic Plots

```
plot(m1)
```


Residuals vs Fitted

lm(log(HPI) ~ log(Personal_Income) + log(Poverty_Percentage) + log(Unemploy ...

Normal Q–Q

Theoretical Quantiles
lm(log(HPI) ~ log(Personal_Income) + log(Poverty_Percentage) + log(Unemploy ...

Scale−Location

Fitted values
lm(log(HPI) ~ log(Personal_Income) + log(Poverty_Percentage) + log(Unemploy ...

Residuals vs Leverage

1428
1912
1880

Cook's distance

Leverage
lm(log(HPI) ~ log(Personal_Income) + log(Poverty_Percentage) + log(Unemploy ...