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:2542        Length:2542        Min.   :  82.32   Min.   : 22440
##  Class :character   Class :character   1st Qu.: 182.93   1st Qu.: 38296
##  Mode  :character   Mode  :character   Median : 237.03   Median : 43505
##                                        Mean   : 307.10   Mean   : 45912
##                                        3rd Qu.: 359.24   3rd Qu.: 50337
##                                        Max.   :2266.07   Max.   :229825
##  Poverty_Percentage   Population        HighSchoolLess  HighSchoolOnly
##  Min.   : 2.70      Min.   :     728   Min.   : 1.50   Min.   : 7.80
##  1st Qu.:10.10      1st Qu.:   15233   1st Qu.: 8.30   1st Qu.:29.50
##  Median :13.00      Median :   31638   Median :11.40   Median :34.40
##  Mean   :13.78      Mean   :  118076   Mean   :12.51   Mean   :33.99
##  3rd Qu.:16.60      3rd Qu.:   79796   3rd Qu.:15.80   3rd Qu.:38.90
##  Max.   :38.20      Max.   :10039107   Max.   :46.70   Max.   :54.50
##   SomeCollege    BachelorAndHigher Unemployment_Rate
##  Min.   :11.20   Min.   : 7.20     Min.   : 1.600
##  1st Qu.:27.70   1st Qu.:15.80     1st Qu.: 3.100
##  Median :31.00   Median :20.00     Median : 3.700
##  Mean   :31.02   Mean   :22.48     Mean   : 3.925
##  3rd Qu.:34.20   3rd Qu.:26.80     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

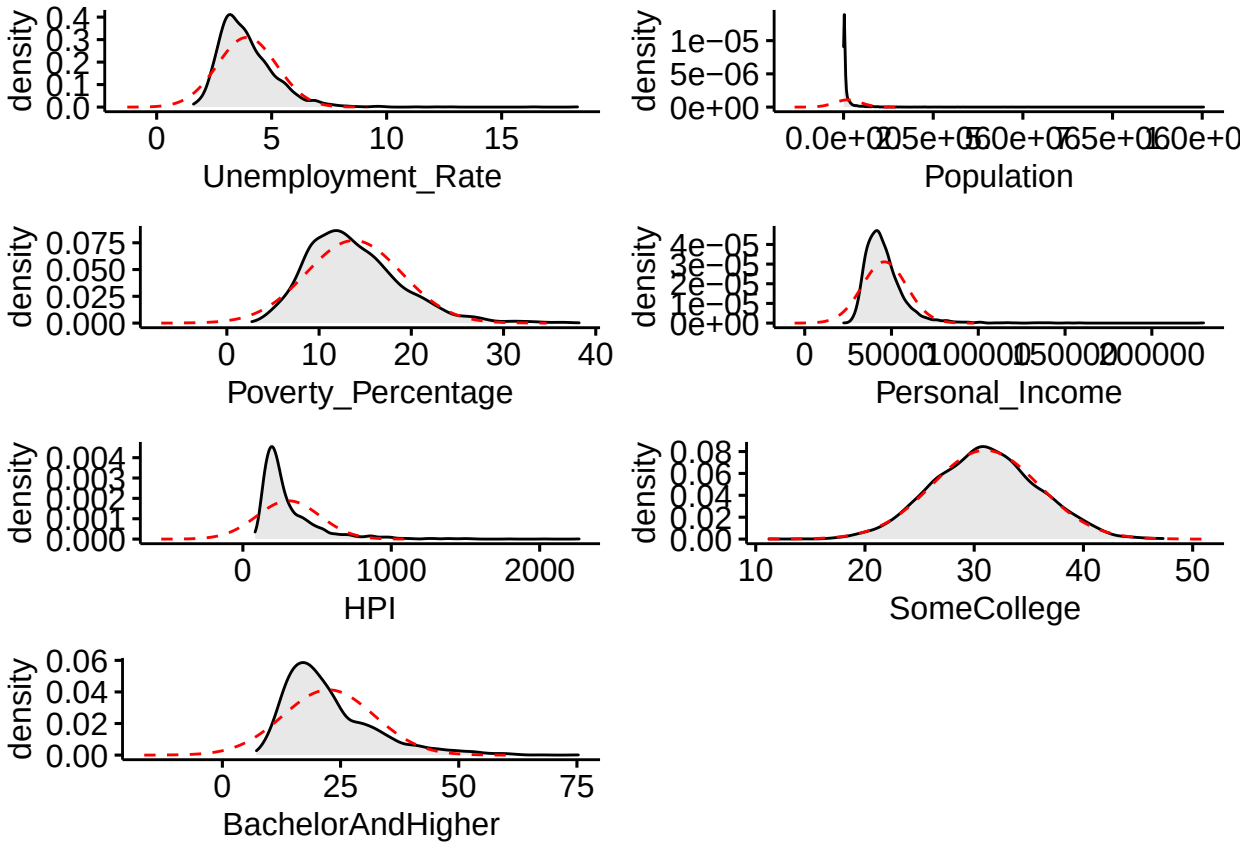
```
##                   HighSchoolLess HighSchoolOnly SomeCollege BachelorAndHigher
## HighSchoolLess         1.0000000      0.2816171 -0.39276930       -0.60303055
## HighSchoolOnly         0.2816171      1.0000000 -0.25170353       -0.79331390
## SomeCollege           -0.3927693     -0.2517035  1.00000000       -0.08764132
## BachelorAndHigher     -0.6030305     -0.7933139 -0.08764132        1.00000000
```

Histogram

```
library(ggpubr)
```

```
## Loading required package: ggplot2
```

```
a<-ggdensity(dt, x = "Unemployment_Rate", fill = "lightgray") +
  stat_overlay_normal_density(color = "red", linetype = "dashed")
b<-ggdensity(dt, x = "Population", fill = "lightgray") +
  stat_overlay_normal_density(color = "red", linetype = "dashed")
c<-ggdensity(dt, x = "Poverty_Percentage", fill = "lightgray") +
  stat_overlay_normal_density(color = "red", linetype = "dashed")
d<-ggdensity(dt, x = "Personal_Income", fill = "lightgray") +
  stat_overlay_normal_density(color = "red", linetype = "dashed")
e<-ggdensity(dt, x = "HPI", fill = "lightgray") +
  stat_overlay_normal_density(color = "red", linetype = "dashed")
f<-ggdensity(dt, x = "SomeCollege", fill = "lightgray") +
  stat_overlay_normal_density(color = "red", linetype = "dashed")
g<-ggdensity(dt, x = "BachelorAndHigher", fill = "lightgray") +
  stat_overlay_normal_density(color = "red", linetype = "dashed")
ggarrange(a,b,c,d,e,f,g, ncol = 2, nrow = 4)
```



Histogram for logtransformation

```
temp=dt
temp$HPI <- log(dt$HPI)
temp$Personal_Income <- log(dt$Personal_Income)
temp$Poverty_Percentage <- log(dt$Poverty_Percentage)
temp$Population <- log(dt$Population)
temp$HighSchoolLess <- log(dt$HighSchoolLess)
temp$BachelorAndHigher <- log(dt$BachelorAndHigher)
temp$Unemployment_Rate <- log(dt$Unemployment_Rate)
```

```
library(ggpubr)
a<-ggdensity(temp, x = "Unemployment_Rate", fill = "lightgray") +
  stat_overlay_normal_density(color = "red", linetype = "dashed")
b<-ggdensity(temp, x = "Population", fill = "lightgray") +
  stat_overlay_normal_density(color = "red", linetype = "dashed")
c<-ggdensity(temp, x = "Poverty_Percentage", fill = "lightgray") +
  stat_overlay_normal_density(color = "red", linetype = "dashed")
d<-ggdensity(temp, x = "Personal_Income", fill = "lightgray") +
  stat_overlay_normal_density(color = "red", linetype = "dashed")
e<-ggdensity(temp, x = "HPI", fill = "lightgray") +
  stat_overlay_normal_density(color = "red", linetype = "dashed")
f<-ggdensity(temp, x = "SomeCollege", fill = "lightgray") +
  stat_overlay_normal_density(color = "red", linetype = "dashed")
g<-ggdensity(temp, x = "BachelorAndHigher", fill = "lightgray") +
  stat_overlay_normal_density(color = "red", linetype = "dashed")
ggarrange(a,b,c,d,e,f,g, ncol = 2, nrow = 4)
```

Model fitting

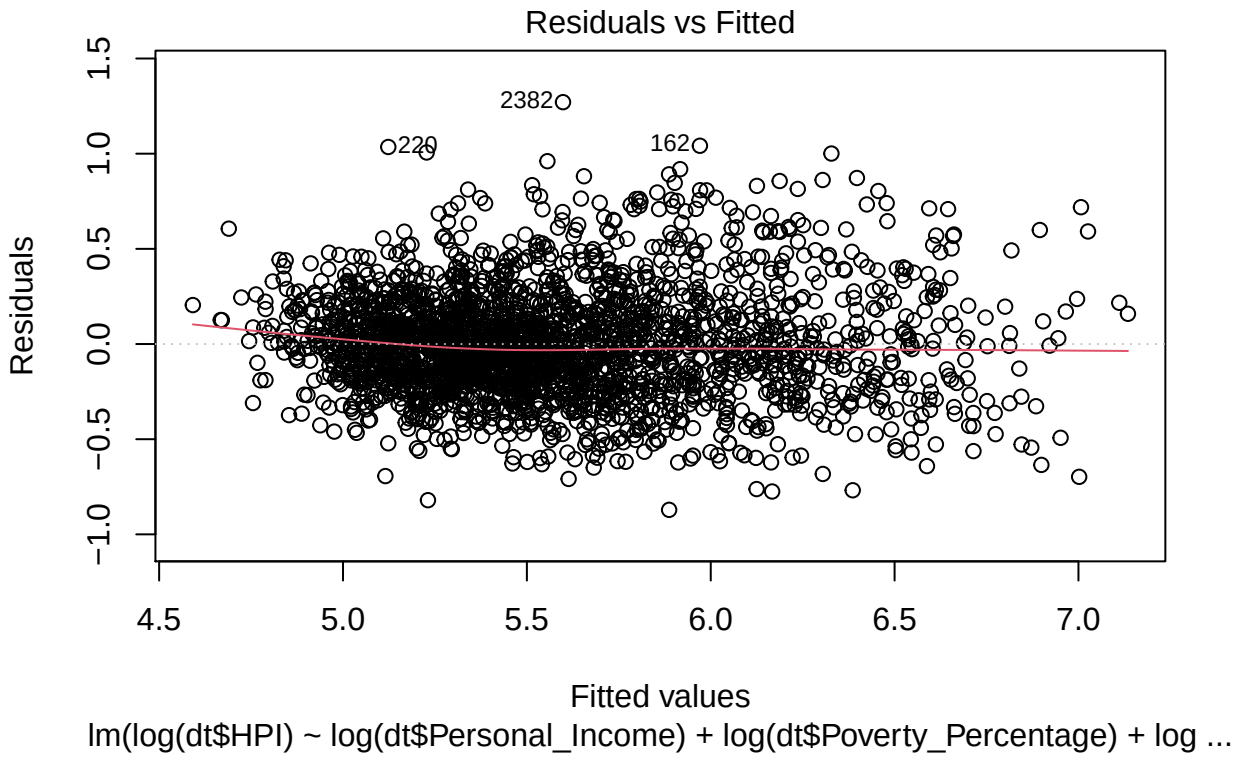
```
m1 = lm(log(dt$HPI)~log(dt$Personal_Income)+log(dt$Poverty_Percentage)+log(dt$Unemployment_Rate)+log(dt$
summary(m1)
```

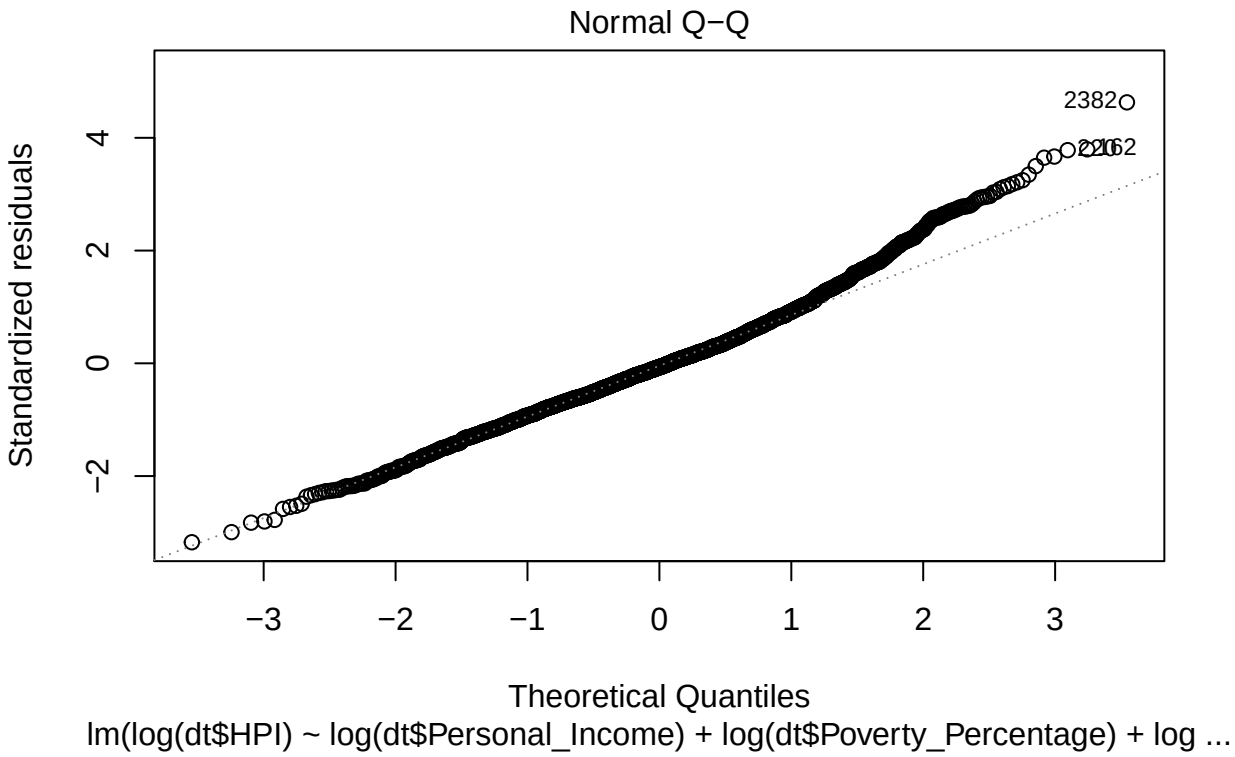
```
##
## Call:
## lm(formula = log(dt$HPI) ~ log(dt$Personal_Income) + log(dt$Poverty_Percentage) +
##     log(dt$Unemployment_Rate) + log(dt$Population) + dt$SomeCollege +
##     log(dt$BachelorAndHigher))
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.87114 -0.17879 -0.01863  0.15435  1.27066
##
## Coefficients:
##                               Estimate Std. Error t value Pr(>|t|)
## (Intercept)                 -3.747673   0.442042  -8.478  < 2e-16 ***
## log(dt$Personal_Income)      0.555545   0.040071  13.864  < 2e-16 ***
## log(dt$Poverty_Percentage) -0.073409   0.022199  -3.307 0.000957 ***
## log(dt$Unemployment_Rate)   0.024234   0.022164   1.093 0.274312
```

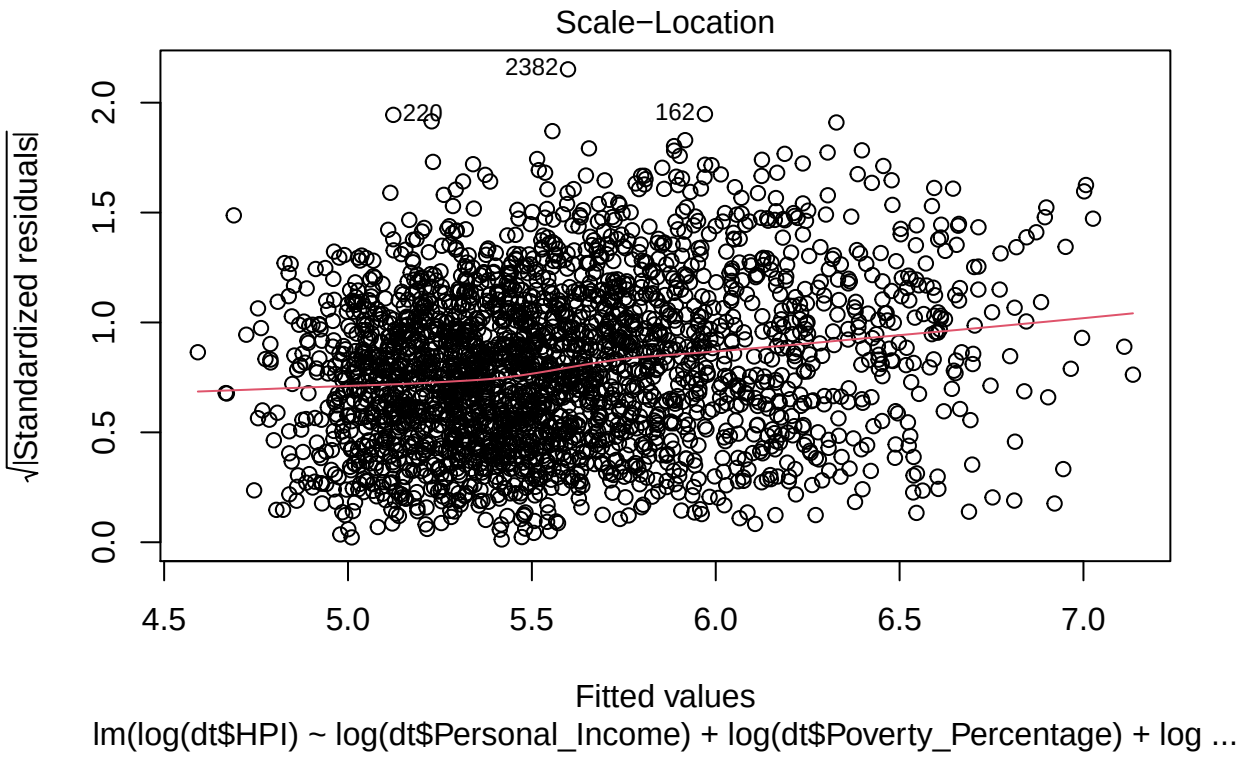
```
## log(dt$Population)          0.237839   0.004941  48.136  < 2e-16 ***
## dt$SomeCollege              0.012988   0.001178  11.021  < 2e-16 ***
## log(dt$BachelorAndHigher)   0.203184   0.022838   8.897  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2748 on 2535 degrees of freedom
## Multiple R-squared:  0.7169, Adjusted R-squared:  0.7162
## F-statistic:  1070 on 6 and 2535 DF,  p-value: < 2.2e-16
```

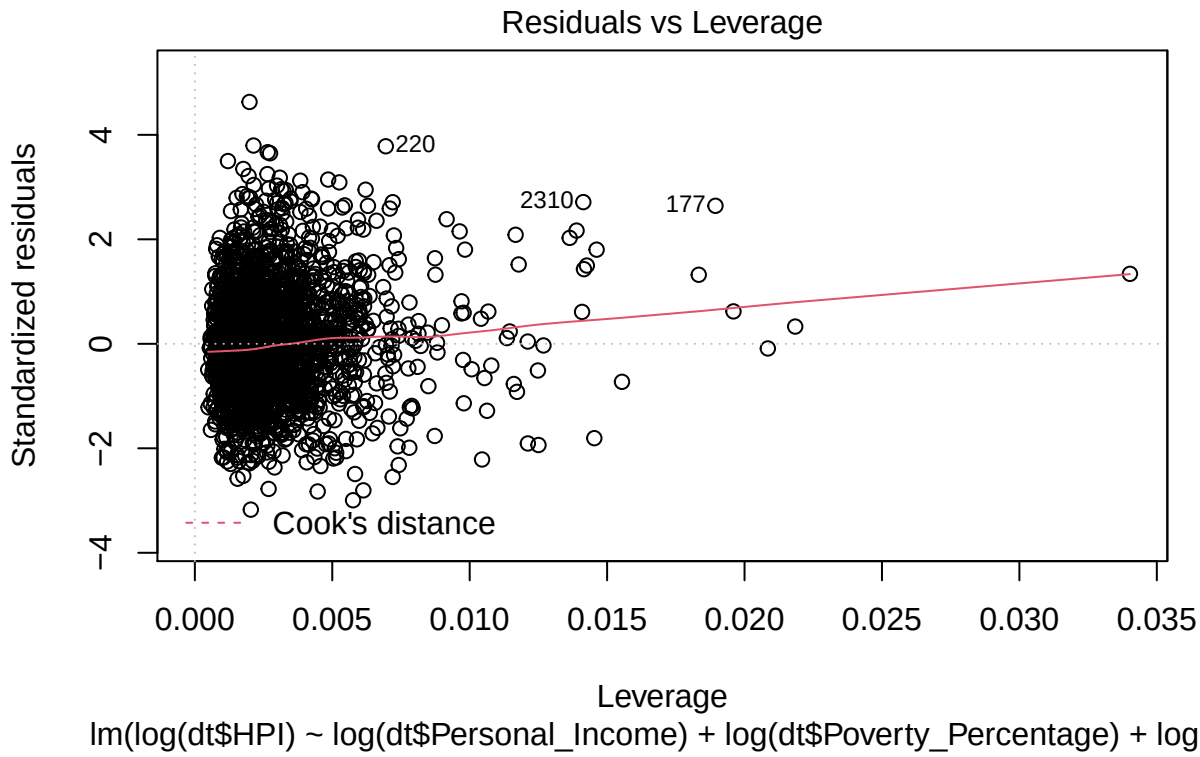
Diagnostic Plots

```
plot(m1)
```




Normal Q−Q
Standardized residuals
Theoretical Quantiles
lm(log(dt$HPI) ~ log(dt$Personal_Income) + log(dt$Poverty_Percentage) + log ...
2382
2206
2062
−3
−2
−1
0
1
2
3
−2
0
2
4


Scale−Location
2382
220
162
√|Standardized residuals|
0.0
0.5
1.0
1.5
2.0
4.5
5.0
5.5
6.0
6.5
7.0
Fitted values
lm(log(dt$HPI) ~ log(dt$Personal_Income) + log(dt$Poverty_Percentage) + log ...




```
car::vif(m1)
```

```
##    log(dt$Personal_Income) log(dt$Poverty_Percentage)
##                   2.772289                   2.304134
##  log(dt$Unemployment_Rate)         log(dt$Population)
##                   1.416236                   1.444652
##             dt$SomeCollege  log(dt$BachelorAndHigher)
##                   1.094225                   2.627954
```