

R Notebook

```
library(lmtest)
```

```
## Loading required package: zoo
```

```
##
```

```
## Attaching package: 'zoo'
```

```
## The following objects are masked from 'package:base':
```

```
##
```

```
##      as.Date, as.Date.numeric
```

```
library(sandwich)
```

```
rm(list = ls())
```

```
lm + coding self
```

```
all in function
```

```
dataset <- read.csv("/Users/chenpoming/Documents/GitHub/Quantitative Method/Quantitative Method/data_for")
```

```
dat <- dataset
```

```
miles <- dat$MILES
```

```
income <- dat$INCOME
```

```
age <- dat$AGE
```

```
kids <- dat$KIDS
```

```
lse <- lm(miles~income+age+kids)
```

```
summary(lse)
```

```
##
```

```
## Call:
```

```
## lm(formula = miles ~ income + age + kids)
```

```
##
```

```
## Residuals:
```

```
##      Min       1Q   Median       3Q      Max  
## -1198.14 -295.31   17.98   287.54  1549.41
```

```
##
```

```
## Coefficients:
```

```
##              Estimate Std. Error t value Pr(>|t|)  
## (Intercept) -391.548    169.775  -2.306  0.0221 *  
## income       14.201      1.800   7.889 2.10e-13 ***  
## age          15.741      3.757   4.189 4.23e-05 ***  
## kids        -81.826     27.130  -3.016  0.0029 **
```

```
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##
```

```
## Residual standard error: 452.3 on 196 degrees of freedom
```

```
## Multiple R-squared:  0.3406, Adjusted R-squared:  0.3305
```

```
## F-statistic: 33.75 on 3 and 196 DF,  p-value: < 2.2e-16
```

```
y <- as.matrix(miles)
```

```
X <- cbind(1,income,age,kids)
```

```
b <- crossprod(solve(crossprod(X,X)),crossprod(X,y))
```

```
e <- y-X%*%b
esq <- e^2

# number of row
n <- nrow(X)
# number of column
k <- ncol(X)
# degree of freedom
nmk <- n-k
```

BP test

```
#
#R square is the focus
#without intercept we will get adjusted Rsquare
BP <- lm(esq~income+age+kids)
summary(BP)

##
## Call:
## lm(formula = esq ~ income + age + kids)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -496114 -161313  -51681   69301 1944762
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    64235     111224   0.578  0.56424
## income           6194         1179   5.252  3.9e-07 ***
## age            -7513         2462  -3.052  0.00259 **
## kids           37256        17773   2.096  0.03735 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 296300 on 196 degrees of freedom
## Multiple R-squared:  0.1451, Adjusted R-squared:  0.132
## F-statistic: 11.09 on 3 and 196 DF,  p-value: 9.37e-07
LMBP <- n*0.1451
LMBP

## [1] 29.02

#29.02      (      H0 Homosterticity)

# p-value
pvchisq <- 1-pchisq(LMBP,df=3)
pvchisq

## [1] 2.217863e-06
```

White test

```

#
# W      ()
# Rsquare
# LMW <- n*0.2
# LMW=
incomesq <- income^2
agesq <- age^2
kidssq <- kids^2
income_age <- income*age
age_kids <- age*kids
kids_income <- kids*income

W <- lm(esq~income+age+kids+incomesq+agesq+kidssq+income_age+age_kids+kids_income)
summary(W)

```

```

##
## Call:
## lm(formula = esq ~ income + age + kids + incomesq + agesq + kidssq +
##      income_age + age_kids + kids_income)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -570705 -150182  -48644   64674 1698102
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -630678.39   523380.42  -1.205   0.22970
## income       21644.67    7831.17    2.764   0.00627 **
## age          1529.98    25030.70    0.061   0.95132
## kids         81985.66   119500.90    0.686   0.49351
## incomesq      -44.33      50.26   -0.882   0.37892
## agesq         220.72     329.28    0.670   0.50347
## kidssq       14459.40    15165.47    0.953   0.34158
## income_age   -315.05     131.56   -2.395   0.01760 *
## age_kids     -5269.67    2702.37   -1.950   0.05265 .
## kids_income   2101.33     923.23    2.276   0.02396 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 291100 on 190 degrees of freedom
## Multiple R-squared:  0.2, Adjusted R-squared:  0.1621
## F-statistic: 5.277 on 9 and 190 DF, p-value: 2.022e-06

```

```

LMW <- n*0.2
LMW

```

```
## [1] 40
```

```

pvchisq <- 1-pchisq(LMW,df=9)
pvchisq

```

```
## [1] 7.598525e-06
```

```
# 7.598525e-06 0 BP White test
```

```
function      BP test
Star <- array(0,dim=c(k,k))
class(Star)

## [1] "matrix"
?array

for(i in 1:n){
  x <- as.matrix(X[i,])
  iter <- (n/nmk)*esq[i]*x%*%t(x)
  Star <- Star+iter
}

AsyVCOV <- solve(crossprod(X,X))%*%Star%*%solve(crossprod(X,X))

Asyvar <- as.matrix(diag(AsyVCOV))

Asyse <- Asyvar^0.5

trat <- b/Asyse

pvt <- 2*pt(-abs(trat),df=nmk)

results <- cbind(b,Asyse,trat,pvt)
colnames(results) <- c("Estimate","Std. Error","t value","Pr(>|t|)")
rownames(results) <- c("intercept","income","age","kids")
round(results,digits=4)

##           Estimate Std. Error t value Pr(>|t|)
## intercept -391.5480    142.6548  -2.7447  0.0066
## income      14.2013     1.9389   7.3246  0.0000
## age         15.7409     3.9657   3.9692  0.0001
## kids        -81.8264    29.1544  -2.8067  0.0055

miles_s <- miles/income^0.5
intercept_s <- 1/income^0.5
income_s <- income/income^0.5
age_s <- age/income^0.5
kids_s <- kids/income^0.5

wlse <- lm(miles_s~intercept_s+income_s+age_s+kids_s-1)
summary(wlse)

##
## Call:
## lm(formula = miles_s ~ intercept_s + income_s + age_s + kids_s -
##      1)
```

```
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -131.336  -37.983    2.073   35.993  169.751
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## intercept_s -408.374    145.717  -2.803  0.00558 **
## income_s      13.971     1.648   8.476 5.54e-15 ***
## age_s         16.348     3.422   4.777 3.48e-06 ***
## kids_s        -78.363    24.736  -3.168  0.00178 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 54.43 on 196 degrees of freedom
## Multiple R-squared:  0.8595, Adjusted R-squared:  0.8567
## F-statistic: 299.8 on 4 and 196 DF,  p-value: < 2.2e-16
```

```
BPtest <- bptest(lse)
```

```
BPtest
```

```
##
## studentized Breusch-Pagan test
##
## data:  lse
## BP = 29.015, df = 3, p-value = 2.223e-06
```

```
Wtest <- bptest(lse, ~income+age+kids+incomesq+agesq+kidssq+income_age+age_kids+kids_income)
Wtest
```

```
##
## studentized Breusch-Pagan test
##
## data:  lse
## BP = 39.993, df = 9, p-value = 7.621e-06
```

```
HCVCOV <- vcovHC(lse)
```

```
HCresults <- coeftest(lse,vcov=HCVCOV)
```

```
HCresults
```

```
##
## t test of coefficients:
##
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -391.5480    144.3576  -2.7123 0.0072750 **
## income       14.2013     1.9850   7.1542 1.639e-11 ***
## age          15.7409     4.0346   3.9015 0.0001313 ***
## kids        -81.8264    29.6644  -2.7584 0.0063582 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
wls <- lm(miles~income+age+kids,weights=1/income)
summary(wls)
```

```
##
## Call:
## lm(formula = miles ~ income + age + kids, weights = 1/income)
##
## Weighted Residuals:
##      Min        1Q      Median        3Q        Max
## -131.336  -37.983    2.073   35.993  169.751
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -408.374    145.717  -2.803  0.00558 **
## income       13.971      1.648   8.476 5.54e-15 ***
## age          16.348      3.422   4.777 3.48e-06 ***
## kids         -78.363     24.736  -3.168  0.00178 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 54.43 on 196 degrees of freedom
## Multiple R-squared:  0.3907, Adjusted R-squared:  0.3814
## F-statistic: 41.9 on 3 and 196 DF,  p-value: < 2.2e-16
```