PanelData

R作Panel Data回归不同估计量的语法，数据和主要程序参考Econometrics Academy <https://sites.google.com/site/econometricsacademy/econometrics-models/panel-data-models> # Panel Data Models in R # Copyright 2013 by Ani Katchova

#读入数据,并设置为Panel Data  
mydata<- read.csv("D:/download/panel\_wage.csv") %>%  
 plm.data(index=c("id","t")) %>%  
 tbl\_df %>%  
 print

## Source: local data frame [4,165 x 22]  
##   
## id t exp wks occ ind south smsa ms fem union ed blk lwage tdum1 tdum2  
## 1 1 1 3 32 0 0 1 0 1 0 0 9 0 5.561 1 0  
## 2 1 2 4 43 0 0 1 0 1 0 0 9 0 5.720 0 1  
## 3 1 3 5 40 0 0 1 0 1 0 0 9 0 5.996 0 0  
## 4 1 4 6 39 0 0 1 0 1 0 0 9 0 5.996 0 0  
## 5 1 5 7 42 0 1 1 0 1 0 0 9 0 6.061 0 0  
## 6 1 6 8 35 0 1 1 0 1 0 0 9 0 6.174 0 0  
## 7 1 7 9 32 0 1 1 0 1 0 0 9 0 6.244 0 0  
## 8 2 1 30 34 1 0 0 0 1 0 0 11 0 6.163 1 0  
## 9 2 2 31 27 1 0 0 0 1 0 0 11 0 6.215 0 1  
## 10 2 3 32 33 1 1 0 0 1 0 1 11 0 6.263 0 0  
## .. .. . ... ... ... ... ... ... .. ... ... .. ... ... ... ...  
## Variables not shown: tdum3 (int), tdum4 (int), tdum5 (int), tdum6 (int),  
## tdum7 (int), exp2 (int)

# 描述性统计  
summary(mydata$lwage)

## Min. 1st Qu. Median Mean 3rd Qu. Max.   
## 4.61 6.40 6.68 6.68 6.95 8.54

summary(select(mydata,exp, exp2, wks, ed))

## exp exp2 wks ed   
## Min. : 1.0 Min. : 1 Min. : 5.0 Min. : 4.0   
## 1st Qu.:11.0 1st Qu.: 121 1st Qu.:46.0 1st Qu.:12.0   
## Median :18.0 Median : 324 Median :48.0 Median :12.0   
## Mean :19.9 Mean : 514 Mean :46.8 Mean :12.8   
## 3rd Qu.:29.0 3rd Qu.: 841 3rd Qu.:50.0 3rd Qu.:16.0   
## Max. :51.0 Max. :2601 Max. :52.0 Max. :17.0

# 混合OLS估计量  
pooling <- plm(lwage~exp+exp2+wks+ed, data=mydata, model= "pooling")  
summary(pooling)

## Oneway (individual) effect Pooling Model  
##   
## Call:  
## plm(formula = lwage ~ exp + exp2 + wks + ed, data = mydata, model = "pooling")  
##   
## Balanced Panel: n=595, T=7, N=4165  
##   
## Residuals :  
## Min. 1st Qu. Median 3rd Qu. Max.   
## -2.160000 -0.250000 0.000273 0.268000 2.130000   
##   
## Coefficients :  
## Estimate Std. Error t-value Pr(>|t|)  
## (Intercept) 4.91e+00 6.73e-02 72.89 < 2e-16  
## exp 4.47e-02 2.39e-03 18.67 < 2e-16  
## exp2 -7.16e-04 5.28e-05 -13.56 < 2e-16  
## wks 5.83e-03 1.18e-03 4.93 8.7e-07  
## ed 7.60e-02 2.23e-03 34.15 < 2e-16  
##   
## Total Sum of Squares: 887  
## Residual Sum of Squares: 635  
## R-Squared : 0.284   
## Adj. R-Squared : 0.283   
## F-statistic: 411.624 on 4 and 4160 DF, p-value: <2e-16

# 组间估计量  
between <- plm(lwage~exp+exp2+wks+ed, data=mydata, model= "between")  
summary(between)

## Oneway (individual) effect Between Model  
##   
## Call:  
## plm(formula = lwage ~ exp + exp2 + wks + ed, data = mydata, model = "between")  
##   
## Balanced Panel: n=595, T=7, N=4165  
##   
## Residuals :  
## Min. 1st Qu. Median 3rd Qu. Max.   
## -0.9780 -0.2200 0.0366 0.2500 0.9860   
##   
## Coefficients :  
## Estimate Std. Error t-value Pr(>|t|)  
## (Intercept) 4.683039 0.210099 22.29 < 2e-16  
## exp 0.038153 0.005697 6.70 5.0e-11  
## exp2 -0.000631 0.000126 -5.02 6.8e-07  
## wks 0.013090 0.004066 3.22 0.0014  
## ed 0.073784 0.004898 15.06 < 2e-16  
##   
## Total Sum of Squares: 92.3  
## Residual Sum of Squares: 62.2  
## R-Squared : 0.326   
## Adj. R-Squared : 0.324   
## F-statistic: 71.4768 on 4 and 590 DF, p-value: <2e-16

# 一阶差分估计量  
  
firstdiff <- plm(lwage~exp+exp2+wks+ed, data=mydata, model= "fd")  
summary(firstdiff)

## Oneway (individual) effect First-Difference Model  
##   
## Call:  
## plm(formula = lwage ~ exp + exp2 + wks + ed, data = mydata, model = "fd")  
##   
## Balanced Panel: n=595, T=7, N=4165  
##   
## Residuals :  
## Min. 1st Qu. Median 3rd Qu. Max.   
## -2.11000 -0.06550 -0.00958 0.04840 2.33000   
##   
## Coefficients :  
## Estimate Std. Error t-value Pr(>|t|)  
## (intercept) 0.117065 0.006311 18.55 < 2e-16  
## exp2 -0.000532 0.000139 -3.82 0.00014  
## wks -0.000268 0.000565 -0.47 0.63485  
##   
## Total Sum of Squares: 118  
## Residual Sum of Squares: 118  
## R-Squared : 0.00411   
## Adj. R-Squared : 0.0041   
## F-statistic: 7.35691 on 2 and 3567 DF, p-value: 0.000648

# 固定效应估计量  
fixed <- plm(lwage~exp+exp2+wks+ed, data=mydata, model= "within")  
summary(fixed)

## Oneway (individual) effect Within Model  
##   
## Call:  
## plm(formula = lwage ~ exp + exp2 + wks + ed, data = mydata, model = "within")  
##   
## Balanced Panel: n=595, T=7, N=4165  
##   
## Residuals :  
## Min. 1st Qu. Median 3rd Qu. Max.   
## -1.81000 -0.05110 0.00371 0.06140 1.94000   
##   
## Coefficients :  
## Estimate Std. Error t-value Pr(>|t|)  
## exp 1.14e-01 2.47e-03 46.09 <2e-16  
## exp2 -4.24e-04 5.46e-05 -7.77 1e-14  
## wks 8.36e-04 6.00e-04 1.39 0.16  
##   
## Total Sum of Squares: 241  
## Residual Sum of Squares: 82.6  
## R-Squared : 0.657   
## Adj. R-Squared : 0.562   
## F-statistic: 2273.74 on 3 and 3567 DF, p-value: <2e-16

# 随机效应估计量  
random <- plm(lwage~exp+exp2+wks+ed, data=mydata, model= "random")  
summary(random)

## Oneway (individual) effect Random Effect Model   
## (Swamy-Arora's transformation)  
##   
## Call:  
## plm(formula = lwage ~ exp + exp2 + wks + ed, data = mydata, model = "random")  
##   
## Balanced Panel: n=595, T=7, N=4165  
##   
## Effects:  
## var std.dev share  
## idiosyncratic 0.0232 0.1522 0.18  
## individual 0.1021 0.3195 0.82  
## theta: 0.823   
##   
## Residuals :  
## Min. 1st Qu. Median 3rd Qu. Max.   
## -2.0400 -0.1060 0.0071 0.1150 2.0900   
##   
## Coefficients :  
## Estimate Std. Error t-value Pr(>|t|)  
## (Intercept) 3.83e+00 9.36e-02 40.9 <2e-16  
## exp 8.89e-02 2.82e-03 31.5 <2e-16  
## exp2 -7.73e-04 6.23e-05 -12.4 <2e-16  
## wks 9.66e-04 7.43e-04 1.3 0.19  
## ed 1.12e-01 6.06e-03 18.4 <2e-16  
##   
## Total Sum of Squares: 261  
## Residual Sum of Squares: 151  
## R-Squared : 0.42   
## Adj. R-Squared : 0.419   
## F-statistic: 753.113 on 4 and 4160 DF, p-value: <2e-16

# 随机效应 VS OLS 的LM检验  
plmtest(pooling)

##   
## Lagrange Multiplier Test - (Honda)  
##   
## data: lwage ~ exp + exp2 + wks + ed  
## normal = 72.06, p-value < 2.2e-16  
## alternative hypothesis: significant effects

# 固定效应 VS OLS的LM检验  
pFtest(fixed, pooling)

##   
## F test for individual effects  
##   
## data: lwage ~ exp + exp2 + wks + ed  
## F = 40.24, df1 = 593, df2 = 3567, p-value < 2.2e-16  
## alternative hypothesis: significant effects

# 固定效应 VS 随机效应的Hausman 检验  
phtest(random, fixed)

##   
## Hausman Test  
##   
## data: lwage ~ exp + exp2 + wks + ed  
## chisq = 6191, df = 3, p-value < 2.2e-16  
## alternative hypothesis: one model is inconsistent