April_19_2020_DiD_R_Stata Demo
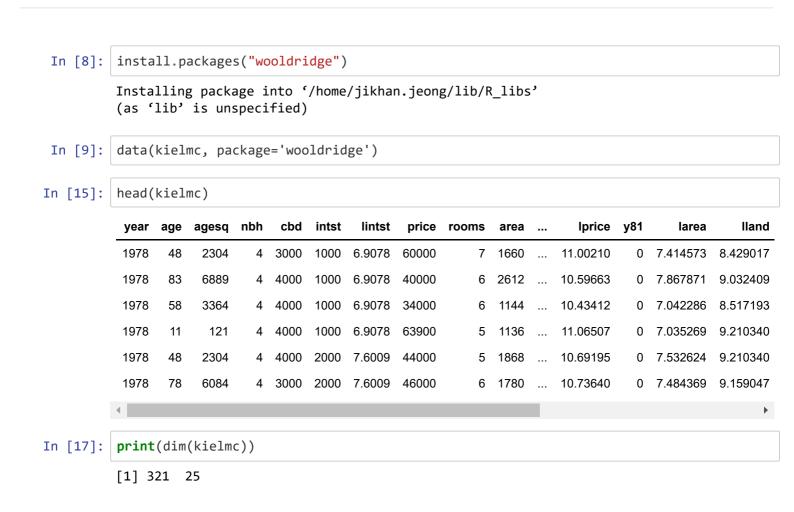
- Name: Jikhan Jeong
- This is for better understanding
- Ref: http://www.urfie.net/index.html (http://www.urfie.net/index.html)

Part 1. Using R: Wooldrige Example 13.3: Effect of a Garbage Incinerator's location on Housing Price

```
In [8]:  install.packages("wooldridge")

         Installing package into '/home/jikhan.jeong/lib/R_libs'
         (as 'lib' is unspecified)
```

```
In [9]:  data(kielmc, package='wooldridge')
```

```
In [15]:  head(kielmc)
```

year	age	agesq	nbh	cbd	intst	lintst	price	rooms	area	...	lprice	y81	larea	lland
1978	48	2304	4	3000	1000	6.9078	60000	7	1660	...	11.00210	0	7.414573	8.429017
1978	83	6889	4	4000	1000	6.9078	40000	6	2612	...	10.59663	0	7.867871	9.032409
1978	58	3364	4	4000	1000	6.9078	34000	6	1144	...	10.43412	0	7.042286	8.517193
1978	11	121	4	4000	1000	6.9078	63900	5	1136	...	11.06507	0	7.035269	9.210340
1978	48	2304	4	4000	2000	7.6009	44000	5	1868	...	10.69195	0	7.532624	9.210340
1978	78	6084	4	3000	2000	7.6009	46000	6	1780	...	10.73640	0	7.484369	9.159047

```
In [17]:  print(dim(kielmc))

          [1] 321  25
```

DiD without covariate

```
In [10]:  # Before treatment with treatment dummy : nearinc
          lm(rprice~nearinc, data=kielmc, subset=(year==1978))
```

```
Call:
lm(formula = rprice ~ nearinc, data = kielmc, subset = (year ==
    1978))

Coefficients:
(Intercept)       nearinc
      82517        -18824
```

```
In [11]:  # After treatment with treatment dummy : nearinc
          lm(rprice~nearinc, data=kielmc, subset=(year==1981))
```

```
Call:
lm(formula = rprice ~ nearinc, data = kielmc, subset = (year ==
    1981))

Coefficients:
(Intercept)       nearinc
     101308        -30688
```

```
In [18]:  # Joint regression including an interaction term
          # lmtest : A collection of tests, data sets, and examples for diagnostic checking in
           linear regression models
          # install.package('lmtest') skip install in here (already in here)
          # Ref: https://cran.r-project.org/web/packages/lmtest/index.html
          library(lmtest)
          coeftest( lm(rprice~nearinc*y81, data=kielmc) )
```

```
Loading required package: zoo

Attaching package: 'zoo'

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

    as.Date, as.Date.numeric


t test of coefficients:

              Estimate Std. Error t value  Pr(>|t|)
(Intercept)   82517.2     2726.9 30.2603 < 2.2e-16 ***
nearinc      -18824.4     4875.3 -3.8612 0.0001368 ***
y81           18790.3     4050.1  4.6395 5.117e-06 ***
nearinc:y81  -11863.9     7456.6 -1.5911 0.1125948
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

DiD with log_real_price of housing

```
In [19]:  DiD      <- lm(log(rprice)~nearinc*y81                       , data=kielmc)
          DiDcontr <- lm(log(rprice)~nearinc*y81+age+I(age^2)+log(intst)+ log(land)+log(area)+r
          ooms+baths, data=kielmc)
```

```
In [21]:  # stargazer is an R package that creates LATEX code, HTML code
          # Ref: https://cran.r-project.org/web/packages/stargazer/vignettes/stargazer.pdf
          install.packages("stargazer")
```

Installing package into '/home/jikhan.jeong/lib/R_libs'
(as 'lib' is unspecified)

```
In [22]:  library(stargazer)
```

Please cite as:

 Hlavac, Marek (2018). stargazer: Well-Formatted Regression and Summary Statistics T
ables.
 R package version 5.2.2. https://CRAN.R-project.org/package=stargazer

```
# DiD with log real price without covariate; Average Treatment Effect (ATE) is not si
gnificant
# DiD with log real price with covariate; ATE is significant -> Log - Lin Model : Int
erpreatation : treatment effect decreases housing price around %13.2
stargazer(DiD,DiDcontr,type="text")
```

```
========================================================================
                              Dependent variable:
                        ----------------------------------------------
                                      log(rprice)
                              (1)                      (2)
------------------------------------------------------------------------
nearinc                     -0.340***                 0.032
                            (0.055)                  (0.047)

y81                          0.193***                 0.162***
                            (0.045)                  (0.028)

age                                                  -0.008***
                                                     (0.001)

I(age2)                                               0.00004***
                                                     (0.00001)

log(intst)                                           -0.061*
                                                     (0.032)

log(land)                                             0.100***
                                                     (0.024)

log(area)                                             0.351***
                                                     (0.051)

rooms                                                 0.047***
                                                     (0.017)

baths                                                 0.094***
                                                     (0.028)

nearinc:y81                 -0.063                   -0.132**
                            (0.083)                  (0.052)

Constant                    11.285***                 7.652***
                            (0.031)                  (0.416)

------------------------------------------------------------------------
Observations                  321                      321
R2                           0.246                    0.733
Adjusted R2                  0.239                    0.724
Residual Std. Error    0.338 (df = 317)          0.204 (df = 310)
F Statistic        34.470*** (df = 3; 317) 84.915*** (df = 10; 310)
========================================================================
Note:                                   *p<0.1; **p<0.05; ***p<0.01
```

Part2 Stata DiD Example

- Cross sectional DiD, (Supplyment for Pandel apporach with two-way fixed effects)
- Treatment : Private tutoring for ITA test in Pullman
- Treated Group : JJ group
- Controal Group : Tanos Group
- Time: 2 period, Before = 2005, After = 2007
- Score: ITA test score
- Data will not be shared

```
In [6]: use"A_data3_5.dta",clear
```

```
In [7]: sum
```

```
    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
         sid |      3,390     3732.53    1928.041         2       6904
        year |      3,390        2006    1.000148      2005       2007
     pri_eng |      3,271    .1210639    .3262514         0          1
   score_eng |      3,232    46.50681      22.6498         4        100
```

```
In [8]: by sid, sort : egen float treat = total(pri_eng)
```

```
In [9]: * Interation term indicates that treatment effect of private tutoring will increase I
        TA scoe 4.5 point higher compared to non-treated
        xi: reg score_eng i.treat*i.year
```

```
i.treat            _Itreat_0-1        (naturally coded; _Itreat_0 omitted)
i.year             _Iyear_2005-2007   (naturally coded; _Iyear_2005 omitted)
i.treat*i.year     _ItreXyea_#_#      (coded as above)
```

```
      Source |       SS           df       MS      Number of obs   =      3,232
-------------+----------------------------------   F(3, 3228)      =      56.82
       Model |  83137.2437          3  27712.4146   Prob > F        =     0.0000
    Residual |  1574408.61       3,228  487.735008   R-squared       =     0.0502
-------------+----------------------------------   Adj R-squared   =     0.0493
       Total |  1657545.85       3,231  513.013262   Root MSE        =     22.085
```

```
------------------------------------------------------------------------------
   score_eng |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
   _Itreat_1 |   9.332955   1.272112     7.34   0.000     6.838727    11.82718
 _Iyear_2007 |  -2.707698    .8911114   -3.04   0.002      -4.4549    -.960497
_ItreXyea_1_2007 |   4.578678     1.82546    2.51   0.012     .9995012    8.157856
        _cons |   45.04872    .6141758   73.35   0.000     43.84451    46.25294
------------------------------------------------------------------------------
```

(Supplyment) Two-way Panel DiD: first differentiated approach

- Considering two-way fixed effect with first diffretiation
- Fixed Panel: $score_{it} = \alpha + \beta D_{it} + u_i + \mu_i + e_{it}$
- First Differentiated to remove unobserved charateristics: $\Delta ITA_s core_i = \beta \Delta D_i + (\mu_{after} - \mu_{before}) + \Delta e_i$

In [11]: `by sid, sort: gen time = _n`

In [12]: `tsset sid time`

```
         panel variable:  sid (strongly balanced)
          time variable:  time, 1 to 2
                  delta:  1 unit
```

In [13]: `* Taking private tuturing of ITA in pullman will increase ITA English score 4.44 points lol`
`reg D.score_eng D.pri_eng`

```
      Source |       SS           df       MS      Number of obs   =     1,484
-------------+----------------------------------   F(1, 1482)      =     14.88
       Model |  5508.66039         1  5508.66039   Prob > F        =    0.0001
    Residual |  548809.321     1,482  370.316681   R-squared       =    0.0099
-------------+----------------------------------   Adj R-squared   =    0.0093
       Total |  554317.981     1,483  373.781511   Root MSE        =    19.244

------------------------------------------------------------------------------
  D.score_eng |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     pri_eng |
         D1. |   4.441488   1.151575     3.86   0.000     2.182598    6.700378
             |
       _cons |   -2.50315   .5773373    -4.34   0.000    -3.635636   -1.370665
------------------------------------------------------------------------------
```

In [14]: `# Ha Ha Ha`

Unknown #command

In [ ]: