

# Homework lab10

Amir Jafarzadeh

08/11/2021

```
setwd ("U:/econometrics/Homework 8 November/Homework- Lab 10")
load("dt_wages(2).RData")
```

```
library (data.table)
```

```
## Warning: package 'data.table' was built under R version 4.1.1
```

```
library(stargazer)
```

```
## Warning: package 'stargazer' was built under R version 4.1.1
```

```
##
```

```
## Please cite as:
```

```
## Hlavac, Marek (2018). stargazer: Well-Formatted Regression and Summary Statistics Tables.
```

```
## R package version 5.2.2. https://CRAN.R-project.org/package=stargazer
```

```
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 4.1.1
```

```
summary(dt.wages)
```

```
##      wage      educ      exper      tenure
##  Min.   : 0.530   Min.   : 0.00   Min.   : 1.00   Min.   : 0.000
## 1st Qu.: 3.330   1st Qu.:12.00   1st Qu.: 5.00   1st Qu.: 0.000
## Median : 4.650   Median :12.00   Median :13.50   Median : 2.000
## Mean   : 5.896   Mean   :12.56   Mean   :17.02   Mean   : 5.105
## 3rd Qu.: 6.880   3rd Qu.:14.00   3rd Qu.:26.00   3rd Qu.: 7.000
## Max.   :24.980   Max.   :18.00   Max.   :51.00   Max.   :44.000
##  nonwhite      female      married      numdep
##  Min.   :0.0000   Min.   :0.0000   Min.   :0.0000   Min.   :0.000
## 1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.000
## Median :0.0000   Median :0.0000   Median :1.0000   Median :1.000
## Mean   :0.1027   Mean   :0.4791   Mean   :0.6084   Mean   :1.044
## 3rd Qu.:0.0000   3rd Qu.:1.0000   3rd Qu.:1.0000   3rd Qu.:2.000
## Max.   :1.0000   Max.   :1.0000   Max.   :1.0000   Max.   :6.000
##      smsa      northcen      south      west
##  Min.   :0.0000   Min.   :0.000   Min.   :0.0000   Min.   :0.0000
## 1st Qu.:0.0000   1st Qu.:0.000   1st Qu.:0.0000   1st Qu.:0.0000
## Median :1.0000   Median :0.000   Median :0.0000   Median :0.0000
## Mean   :0.7224   Mean   :0.251   Mean   :0.3555   Mean   :0.1692
## 3rd Qu.:1.0000   3rd Qu.:0.750   3rd Qu.:1.0000   3rd Qu.:0.0000
## Max.   :1.0000   Max.   :1.000   Max.   :1.0000   Max.   :1.0000
##  construc      ndurman      trcompu      trade
```

```
## Min. :0.00000 Min. :0.0000 Min. :0.00000 Min. :0.0000
## 1st Qu.:0.00000 1st Qu.:0.0000 1st Qu.:0.00000 1st Qu.:0.0000
## Median :0.00000 Median :0.0000 Median :0.00000 Median :0.0000
## Mean :0.04563 Mean :0.1141 Mean :0.04373 Mean :0.2871
## 3rd Qu.:0.00000 3rd Qu.:0.0000 3rd Qu.:0.00000 3rd Qu.:1.0000
## Max. :1.00000 Max. :1.0000 Max. :1.00000 Max. :1.0000
## services profserv profocc clerocc
## Min. :0.0000 Min. :0.0000 Min. :0.0000 Min. :0.0000
## 1st Qu.:0.0000 1st Qu.:0.0000 1st Qu.:0.0000 1st Qu.:0.0000
## Median :0.0000 Median :0.0000 Median :0.0000 Median :0.0000
## Mean :0.1008 Mean :0.2586 Mean :0.3669 Mean :0.1673
## 3rd Qu.:0.0000 3rd Qu.:1.0000 3rd Qu.:1.0000 3rd Qu.:0.0000
## Max. :1.0000 Max. :1.0000 Max. :1.0000 Max. :1.0000
## servocc lwage expersq tenursq
## Min. :0.0000 Min. : -0.6349 Min. : 1.0 Min. : 0.00
## 1st Qu.:0.0000 1st Qu.: 1.2030 1st Qu.: 25.0 1st Qu.: 0.00
## Median :0.0000 Median : 1.5369 Median : 182.5 Median : 4.00
## Mean :0.1407 Mean : 1.6233 Mean : 473.4 Mean : 78.15
## 3rd Qu.:0.0000 3rd Qu.: 1.9286 3rd Qu.: 676.0 3rd Qu.: 49.00
## Max. :1.0000 Max. : 3.2181 Max. :2601.0 Max. :1936.00
```

```
ncol(dt.wages)##showing number of column which is the number of variables
```

```
## [1] 24
```

```
nrow(dt.wages) ## showing the number of rows which is the number of observation in each variable
```

```
## [1] 526
```

*#3. 3.) Let's do a thought experiment: Using the slide-deck UEA\_ecoR2PhD CoreLect\_06 ATENT\_Match\_Stk a*  
*##a) a. Compute a difference-in-means estimator when treatment is "south," and the outcome is wage.*

```
lm.a <- lm(wage ~ south, data=dt.wages)
stargazer(lm.a, type='text')
```

```
##
## =====
##                Dependent variable:
##                -----
##                wage
## -----
## south                -0.790**
##                      (0.335)
##
## Constant              6.177***
##                      (0.200)
## -----
## Observations                526
## R2                        0.011
## Adjusted R2                0.009
## Residual Std. Error      3.677 (df = 524)
## F Statistic              5.564** (df = 1; 524)
## =====
## Note:                *p<0.1; **p<0.05; ***p<0.01
```

```
##b) Now focus on race and gender as control variables (in "x") and run a regression estimation of treatment effects
lm.b <- lm(wage ~ nonwhite+female, data=dt.wages)
stargazer(lm.b, type='text')
```

```
##
## =====
##                      Dependent variable:
##                      -----
##                      wage
## -----
## nonwhite              -0.513
##                      (0.499)
##
## female                -2.515***
##                      (0.303)
##
## Constant              7.154***
##                      (0.217)
##
## -----
## Observations          526
## R2                    0.117
## Adjusted R2           0.114
## Residual Std. Error   3.476 (df = 523)
## F Statistic           34.801*** (df = 2; 523)
## =====
## Note:                  *p<0.1; **p<0.05; ***p<0.01
```

##c) Now try to estimate the regression and account for potentially heterogeneous treatment effects

```
lm.c <- lm(wage ~ south + female + nonwhite + south*female + south*nonwhite, data=dt.wages)
stargazer(lm.c, type='text')
```

```
##
## =====
##                      Dependent variable:
##                      -----
##                      wage
## -----
## south                 -1.288***
##                      (0.447)
##
## female                -2.953***
##                      (0.374)
##
## nonwhite              0.155
##                      (0.691)
##
## south:female          1.117*
##                      (0.630)
##
## south:nonwhite        -1.047
##                      (0.996)
##
```

```
## Constant          7.628***
##                   (0.269)
## -----
## Observations      526
## R2                 0.138
## Adjusted R2       0.129
## Residual Std. Error 3.446 (df = 520)
## F Statistic       16.591*** (df = 5; 520)
## =====
## Note:              *p<0.1; **p<0.05; ***p<0.01
```

*##d) Next, try to implement a 2-step fitted regression.*

```
library("ivreg")
```

```
## Warning: package 'ivreg' was built under R version 4.1.1
```

```
ivB <- ivreg(wage~south + female + nonwhite + south*female + south*nonwhite|northcen + female + nonwhite)
summary(ivB)
```

```
##
## Call:
## ivreg(formula = wage ~ south + female + nonwhite + south * female +
##       south * nonwhite | northcen + female + nonwhite + northcen *
##       female + northcen * nonwhite, data = dt.wages)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -5.8143 -2.0670 -0.9409  1.4123 17.9198
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    7.0602    0.4463  15.819 < 2e-16 ***
## south           0.2541    1.0793   0.235  0.814
## female        -2.6193    0.6188  -4.233 2.73e-05 ***
## nonwhite       -0.9891    1.2966  -0.763  0.446
## south:female    0.3719    1.5329   0.243  0.808
## south:nonwhite  0.8293    2.4723   0.335  0.737
##
## Diagnostic tests:
##              df1 df2 statistic p-value
## Weak instruments (south)      3 520   38.626 <2e-16 ***
## Weak instruments (south:female) 3 520   35.962 <2e-16 ***
## Weak instruments (south:nonwhite) 3 520   37.427 <2e-16 ***
## Wu-Hausman      3 517    1.993  0.114
## Sargan          0 NA        NA      NA
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.527 on 520 degrees of freedom
## Multiple R-Squared: 0.09642, Adjusted R-squared: 0.08774
## Wald test: 13.64 on 5 and 520 DF, p-value: 1.601e-12
```

```
stargazer(ivB, type="text")
```

```
##
## =====
##                               Dependent variable:
##                               -----
##                               wage
## -----
## south                        0.254
##                               (1.079)
##
## female                       -2.619***
##                               (0.619)
##
## nonwhite                     -0.989
##                               (1.297)
##
## south:female                 0.372
##                               (1.533)
##
## south:nonwhite              0.829
##                               (2.472)
##
## Constant                     7.060***
##                               (0.446)
##
## -----
## Observations                 526
## R2                           0.096
## Adjusted R2                  0.088
## Residual Std. Error         3.527 (df = 520)
## =====
## Note:                        *p<0.1; **p<0.05; ***p<0.01
```

*##e) Next, consider that you also want to control for experience.*

*###i. Can you do a regression estimation of treatment effects?*

```
lm.e.i <- lm(wage ~ south + female + nonwhite+exper , data=dt.wages)
stargazer(lm.e.i, type='text')
```

```
##
## =====
##                               Dependent variable:
##                               -----
##                               wage
## -----
## south                        -0.923***
##                               (0.315)
##
## female                       -2.521***
##                               (0.300)
##
## nonwhite                     -0.383
##                               (0.496)
##
```

```
## exper                0.029***
##                    (0.011)
##
## Constant            6.985***
##                    (0.307)
##
## -----
## Observations        526
## R2                  0.141
## Adjusted R2         0.135
## Residual Std. Error 3.435 (df = 521)
## F Statistic         21.462*** (df = 4; 521)
## =====
## Note:                *p<0.1; **p<0.05; ***p<0.01

###ii. Can you account for heterogeneous treatment effects w.r.t. experience?
lm.e.ii <- lm(wage ~ south + female + nonwhite + south*exper + exper, data=dt.wages)
stargazer(lm.e.i, type='text')
```

```
##
## =====
##                               Dependent variable:
##                               -----
##                               wage
## -----
## south                      -0.923***
##                          (0.315)
##
## female                     -2.521***
##                          (0.300)
##
## nonwhite                   -0.383
##                          (0.496)
##
## exper                      0.029***
##                          (0.011)
##
## Constant                   6.985***
##                          (0.307)
##
## -----
## Observations              526
## R2                        0.141
## Adjusted R2               0.135
## Residual Std. Error       3.435 (df = 521)
## F Statistic               21.462*** (df = 4; 521)
## =====
## Note:                     *p<0.1; **p<0.05; ***p<0.01
```

```
ivB <- ivreg(wage~south + exper + nonwhite + south*exper |northcen + exper + nonwhite + northcen)
summary(ivB)
```

```
##
## Call:
## ivreg(formula = wage ~ south + exper + nonwhite + south * exper |
```

```
##      northcen + exper + nonwhite + northcen * exper, data = dt.wages)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -4.850 -2.429 -1.087   1.262  18.568
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   4.76070    0.49912   9.538  <2e-16 ***
## south         2.11396    1.35755   1.557   0.1200
## exper         0.05693    0.02220   2.565   0.0106 *
## nonwhite     -0.54795    0.54724  -1.001   0.3171
## south:exper  -0.08292    0.05747  -1.443   0.1497
##
## Diagnostic tests:
##                                df1 df2 statistic p-value
## Weak instruments (south)        2 521    58.304  <2e-16 ***
## Weak instruments (south:exper)  2 521    69.881  <2e-16 ***
## Wu-Hausman                     2 519     2.182    0.114
## Sargan                          0 NA         NA      NA
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.703 on 521 degrees of freedom
## Multiple R-Squared:  0.002171,    Adjusted R-squared: -0.00549
## Wald test: 2.504 on 4 and 521 DF,  p-value: 0.04144
```

```
stargazer(ivB, type="text")
```

```
##
## =====
##                      Dependent variable:
##                      -----
##                      wage
## -----
## south                2.114
##                      (1.358)
##
## exper                0.057**
##                      (0.022)
##
## nonwhite            -0.548
##                      (0.547)
##
## south:exper         -0.083
##                      (0.057)
##
## Constant            4.761***
##                      (0.499)
##
## -----
## Observations                526
## R2                          0.002
## Adjusted R2                 -0.005
## Residual Std. Error      3.703 (df = 521)
```

```
## =====
## Note:                *p<0.1; **p<0.05; ***p<0.01
```

*#4.) Lastly, let's worry that "south" is potentially not a great randomly assigned treatment.  
## a. Provide up to 3 reasons, why this view might be flawed.*

```
stargazer(dt.wages, type="text")
```

```
##
## =====
## Statistic  N    Mean    St. Dev.  Min   Pctl(25) Pctl(75)  Max
## -----
## wage      526  5.896    3.693    0.530   3.330    6.880    24.980
## educ      526 12.563    2.769     0        12       14       18
## exper     526 17.017   13.572    1         5       26       51
## tenure    526  5.105    7.224     0         0        7       44
## nonwhite   526  0.103    0.304     0         0        0        1
## female    526  0.479    0.500     0         0        1        1
## married   526  0.608    0.489     0         0        1        1
## numdep    526  1.044    1.262     0         0        2        6
## smsa      526  0.722    0.448     0         0        1        1
## northcen  526  0.251    0.434     0         0        0.8      1
## south     526  0.356    0.479     0         0        1        1
## west      526  0.169    0.375     0         0        0        1
## construc  526  0.046    0.209     0         0        0        1
## ndurman   526  0.114    0.318     0         0        0        1
## trcompu   526  0.044    0.205     0         0        0        1
## trade     526  0.287    0.453     0         0        1        1
## services  526  0.101    0.301     0         0        0        1
## profserv  526  0.259    0.438     0         0        1        1
## profocc   526  0.367    0.482     0         0        1        1
## clerocc   526  0.167    0.374     0         0        0        1
## servocc   526  0.141    0.348     0         0        0        1
## lwage     526  1.623    0.532   -0.635   1.203    1.929    3.218
## expersq   526 473.435  616.045    1        25       676     2,601
## tenursq   526 78.150   199.435    0         0        49     1,936
## -----
```

```
mean.south <- mean(dt.wages$south)
```

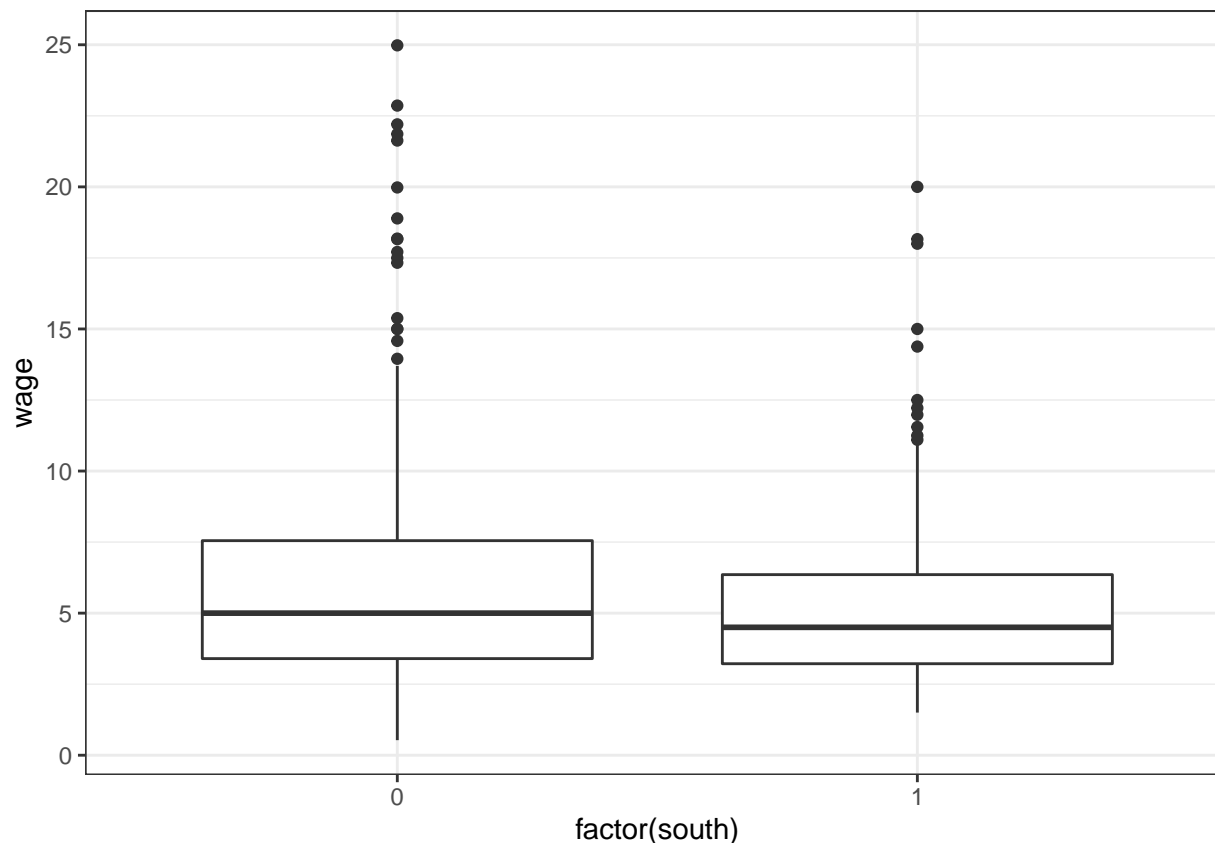
```
mean.south
```

```
## [1] 0.3555133
```

*## b. Provide up to 3 reasons, why it might be justified. (sketch bullets, don't get philosophical)*

```
qplot(data=dt.wages, x=factor(south), y=wage, geom='boxplot') + theme_bw()
```





```
dt.wages[, t.test(wage~south)]
```

```
##
## Welch Two Sample t-test
##
## data: wage by south
## t = 2.5291, df = 464.56, p-value = 0.01177
## alternative hypothesis: true difference in means between group 0 and group 1 is not equal to 0
## 95 percent confidence interval:
##  0.1761943 1.4039912
## sample estimates:
## mean in group 0 mean in group 1
##      6.176991      5.386898
```

##c. Now, let's check the balance in the covariates:

###i. Provide summary statistics for both groups (south=0, south =1) separately. Are the cova  
stargazer(dt.wages [south==0], type = "text")

```
##
## =====
## Statistic  N   Mean   St. Dev.  Min   Pctl(25) Pctl(75)  Max
## -----
## wage      339  6.177   3.960    0.530   3.400    7.550   24.980
## educ      339 12.802   2.556     0      12      14     18
## exper     339 16.510  13.846     1      5      26     49
## tenure    339  5.239   7.429     0      0       7     44
## nonwhite   339  0.080   0.271     0      0       0      1
```

```
## female      339  0.496  0.501    0    0    1    1
## married     339  0.581  0.494    0    0    1    1
## numdep      339  0.962  1.197    0    0    2    5
## smsa        339  0.799  0.401    0    1    1    1
## northcen    339  0.389  0.488    0    0    1    1
## south       339  0.000  0.000    0    0    0    0
## west        339  0.263  0.441    0    0    1    1
## construc    339  0.041  0.199    0    0    0    1
## ndurman     339  0.086  0.280    0    0    0    1
## trcommpu    339  0.041  0.199    0    0    0    1
## trade       339  0.277  0.448    0    0    1    1
## services    339  0.100  0.301    0    0    0    1
## profserv    339  0.286  0.453    0    0    1    1
## profocc     339  0.404  0.491    0    0    1    1
## clerocc     339  0.159  0.366    0    0    0    1
## servocc     339  0.147  0.355    0    0    0    1
## lwage       339  1.659  0.556   -0.635  1.224  2.022  3.218
## expersq     339 463.737 620.913    1    25   676  2,401
## tenursq     339 82.472 211.458    0    0    49  1,936
## -----
```

```
stargazer(dt.wages [south==1], type = "text")
```

```
##
## =====
## Statistic  N    Mean    St. Dev.  Min  Pctl(25) Pctl(75)  Max
## -----
## wage      187  5.387    3.099    1.500  3.220    6.355    20.000
## educ      187 12.128    3.079     0      11     14     18
## exper     187 17.936   13.047     1      6     27     51
## tenure    187  4.861    6.851     0      0      6     31
## nonwhite  187  0.144    0.352     0      0      0      1
## female    187  0.449    0.499     0      0      1      1
## married   187  0.658    0.476     0      0      1      1
## numdep    187  1.193    1.362     0      0      2      6
## smsa      187  0.583    0.494     0      0      1      1
## northcen  187  0.000    0.000     0      0      0      0
## south     187  1.000    0.000     1      1      1      1
## west      187  0.000    0.000     0      0      0      0
## construc  187  0.053    0.226     0      0      0      1
## ndurman   187  0.166    0.373     0      0      0      1
## trcommpu  187  0.048    0.215     0      0      0      1
## trade     187  0.305    0.462     0      0      1      1
## services  187  0.102    0.303     0      0      0      1
## profserv  187  0.209    0.407     0      0      0      1
## profocc   187  0.299    0.459     0      0      1      1
## clerocc   187  0.182    0.387     0      0      0      1
## servocc   187  0.128    0.335     0      0      0      1
## lwage     187  1.559    0.479    0.405  1.169    1.849    2.996
## expersq   187 491.016 608.380     1     36    729    2,601
## tenursq   187 70.316 175.826     0      0     36     961
## -----
```

*##d. Describe verbally how would you set up a matching strategy? Which factors (variables) do you*

*####answer: we should test statistics of the all variables in the treated group (souh==1) and i*

```
####i. yest it is the best strategy to match
```