ECON4150 - Introductory Econometrics Seminar 9

Stock and Watson EE13.1

April 28, 2015

Empirical exercise E13.1: Data

- Names.dta: data from a randomized controlled experiment conducted by Marianne Bertrand and Sendhil Mullainathan
- Names contains resume, call-back and employer information for 4,870 fictitious resumes sent in response to employment advertisements in Chicago and Boston in 2001
- The resumes contained information concerning the race of the applicant
- race is not typically included on a resume
- Randomly assigned "white-sounding" names and "African american sounding" names to resumes and sent out as job application behavior.
- Recorded the number of "call backs" from employers

Empirical exercise E11.2: Data

Variable	**description**
first name	applicant's first name
female	=1 if female, =0 otherwise
black	=1 if black sounding name, =0 otherwise
high	=1 high quality resume, =0 otherwise
call_back	=1 if called back, =0 otherwise
chicago	=1 if data from chicago, =0 otherwise

Empirical exercise E11.2: Data

```
clear
set more off

clear all

cd M:\pc\Desktop\courses\introductory_econometrics\seminar_9

use "Names.dta",clear
summ
```

a) first possibility, ttest command

```
ttest call_back, by(black)

Two-sample t test with equal variances
------------------------------------------------------------------------------
   Group |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
       0 |    2435    .0965092    .0059853     .295349    .0847724    .1082461
       1 |    2435    .0644764    .0049781    .2456501    .0547145    .0742382
---------+--------------------------------------------------------------------
combined |    4870    .0804928    .0038988    .2720826    .0728493    .0881363
---------+--------------------------------------------------------------------
    diff |            .0320329     .007785                .0167708    .0472949
------------------------------------------------------------------------------
    diff = mean(0) - mean(1)                                      t =   4.1147
Ho: diff = 0                     degrees of freedom =     4868

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 1.0000         Pr(|T| > |t|) = 0.0000          Pr(T > t) = 0.0000
```

The call back rate for whites is 0.097 while for blacks is 0.064, the difference is 0.032. The resume with black sounding name are called back 3.2 % less than the resumes with white sounding name. The difference is big, resume with black sounding names are called back 1/3 less of the times.

a) second possibility, summ and then calculate IC

```
summ call_back if black==0
    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
   call_back |      2435    .0965092     .295349          0          1

summ call_back if black==1
    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
   call_back |      2435    .0644764    .2456501          0          1
```

The estimated call back rate for whites ($\hat{P}_W$) is 0.097 while for blacks ($\hat{P}_B$) is 0.064. Difference again is 0.032

a) second possibility, summ and then calculate IC

The 95% confidence interval for the difference between $(\hat{P}_W)$ and $(\hat{P}_B)$ is contructed as in equation 3.21 in the book, $(\hat{P}_W - \hat{P}_B) \pm 1.96 SE(\hat{P}_W - \hat{P}_B)$ where $SE(\hat{P}_W - \hat{P}_B) = \sqrt{\frac{s_W^2}{n_W} + \frac{s_B^2}{n_B}}$ and s_W^2, s_B^2 are the sample variances and n_W, n_B are the number of observations of the two groups.

$$CI = 0.032 \pm 1.96 * \sqrt{\frac{0.295^2}{2435} + \frac{0.246^2}{2435}} = [0.017, 0.047] \tag{1}$$

a) third possibility, regression

```
regress call_back black, r

Linear regression                               Number of obs =    4870
                                                F(  1,  4868) =   16.93
                                                Prob > F      =  0.0000
                                                R-squared     =  0.0035
                                                Root MSE      =  .27164

------------------------------------------------------------------------------
             |               Robust
   call_back |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       black |  -.0320329    .007785    -4.11   0.000    -.0472949   -.0167708
       _cons |   .0965092   .0059853    16.12   0.000     .0847753    .1082431
------------------------------------------------------------------------------
```

The difference is statistically significant different from 0 and the 95 % confidence interval is [-.047 -.017].

b)

```
gen femalexblack = female*black
reg call_back black femalexblack, r

Linear regression                               Number of obs =    4870
                                                F(  2,  4867) =    8.80
                                                Prob > F      =  0.0002
                                                R-squared     =  0.0035
                                                Root MSE      =  .27166

------------------------------------------------------------------------------
             |               Robust
   call_back |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       black |  -.0382214   .0116566    -3.28   0.001    -.0610735   -.0153693
femalexblack |     .00799   .0115272     0.69   0.488    -.0146085    .0305886
       _cons |   .0965092   .0059859    16.12   0.000     .0847741    .1082443
------------------------------------------------------------------------------
```

The coefficient on femaleblack = female*black measures the differential effect of being black for female relative to male. The coefficient of the interaction term is not statistically significant different from 0, so that we can reject the hipotesys that the call back differential differs between man and women.

```
 ttest call_back, by(high)

Two-sample t test with equal variances
------------------------------------------------------------------------------
   Group |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
       0 |    2424    .0734323    .0052991    .2608987     .063041    .0838237
       1 |    2446    .0874898    .0057142    .2826092    .0762845     .098695
---------+--------------------------------------------------------------------
combined |    4870    .0804928    .0038988    .2720826    .0728493    .0881363
---------+--------------------------------------------------------------------
    diff |           -.0140574     .007796               -.0293411    .0012262
------------------------------------------------------------------------------
    diff = mean(0) - mean(1)                                      t =  -1.8032
Ho: diff = 0                     degrees of freedom =     4868

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 0.0357         Pr(|T| > |t|) = 0.0714          Pr(T > t) = 0.9643
```

The difference is not statistical significant at a 5% level

c)

```
gen highxblack = high*black
reg call_back black highxblack high
Linear regression                               Number of obs =    4870
                                                F(  3,  4866) =    6.61
                                                Prob > F      =  0.0002
                                                R-squared     =  0.0044
------------------------------------------------------------------------------
             |               Robust
   call_back |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       black |  -.0231023   .0105901    -2.18   0.029    -.0438636    -.002341
  highxblack |  -.0177808   .0155605    -1.14   0.253    -.0482864    .0127248
        high |   .0229478   .0119584     1.92   0.055     -.000496    .0463917
       _cons |   .0849835   .0080133    10.61   0.000     .0692739    .1006931
------------------------------------------------------------------------------
```

- coefficent on high represents the estimated difference in call backs between high and low white sounding name applications. 2.2 %
- coefficents on high+highxblack represents the estimated difference in call backs between high and low black sounding name applications. 0.5%
- difference in the high-quality/low-quality call back difference for white versus african Americans is then 1.7%, the coefficient on highXblack.
- But, the interaction term is not statistical significant, meaning that the difference is not statistical significant.

d)

```
foreach var of varlist  yearsexp honors volunteer military empholes workinschool ///
 email computerskills specialskills eoe manager supervisor secretary offsupport  ///
 salesrep retailsales req expreq comreq educreq compreq orgreq manuf ///
transcom bankreal trade busservice othservice missind chicago high female college call_back {
di "`var'"
ttest `var', by(black)
scalar p_value_`var' = r(p)
scalar mean1_`var' = r(mu_1)
scalar mean0_`var' = r(mu_2)
}

foreach var of varlist  yearsexp honors volunteer military empholes workinschool  ///
email computerskills specialskills eoe manager supervisor secretary offsupport  ///
salesrep retailsales req expreq comreq educreq compreq orgreq manuf ///
transcom bankreal trade busservice othservice missind chicago high female college call_back {
if `var'== yearsexp {
noisily display _column(1) "Variable "  ///
                _column(15)"mean white" ///
                _column(30)"mean black " ///
                _column(45) "difference" _column(60)"p-value diference=0"
 }

noisily display _column(1) "`var'" _column(17) string(round( mean1_`var',0.001)) ///
                _column(32) string(round(mean0_`var',0.001)) ///
                _column(47) string(round(mean1_`var'-mean0_`var',0.001)) ///
                _column(62) string(round(p_value_`var',0.001))
}
```

d)

Variable	mean white	mean black	difference	p-value diference=0
yearsexp	7.856	7.83	.027	.854
honors	.054	.051	.003	.654
volunteer	.409	.414	-.006	.684
military	.092	.102	-.009	.266
empholes	.45	.446	.004	.773
workinschool	.558	.561	-.003	.84
email	.479	.48	-.001	.954
computerskills	.809	.832	-.024	.03
specialskills	.33	.327	.003	.831
eoe	.291	.291	0	1
manager	.152	.152	0	.968
supervisor	.077	.077	0	1
secretary	.333	.333	0	.976
offsupport	.119	.119	0	1
salesrep	.151	.151	0	1
retailsales	.168	.168	0	1
req	.787	.787	0	1
expreq	.435	.435	0	1
comreq	.125	.125	0	1
educreq	.107	.107	0	1

d)

Variable	mean white	mean black	difference	p-value diference=0
compreq	.437	.437	0	.977
orgreq	.073	.073	0	1
manuf	.083	.083	0	1
transcom	.03	.03	0	1
bankreal	.085	.085	0	1
trade	.214	.214	0	1
busservice	.268	.268	0	1
othservice	.155	.155	0	1
missind	.165	.165	0	1
chicago	.555	.555	0	1
high	.502	.502	0	1
female	.764	.775	-.011	.377
college	.716	.723	-.007	.61
call_back	.097	.064	.032	0

There are only two significant differences in the mean values: the call-back rate (the variable of interest) and computer skills (for which blacknamed resumes had a slightly higher fraction that white-named resumes). Thus, there is no evidence of non-random assignment.

d) F statistic regression

```
reg black  yearsexp honors volunteer military empholes workinschool email computerskills\\\
    specialskills eoe manager supervisor secretary offsupport salesrep retailsales req expreq\\\
    comreq educreq  compreq orgreq manuf transcom bankreal trade busservice othservice\\\
    missind chicago high female college

      Source |       SS       df       MS              Number of obs =    4870
-------------+------------------------------           F( 31,  4838) =    0.29
       Model |  2.24727714    31  .072492811            Prob > F      =  1.0000
    Residual |  1215.25272  4838   .25118907            R-squared     =  0.0018
-------------+------------------------------           Adj R-squared = -0.0045
       Total |      1217.5  4869  .250051345            Root MSE      =  .50119

------------------------------------------------------------------------------
       black |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     yearsexp |   .0007854   .0016711     0.47   0.638    -.0024906    .0040615
       honors |  -.0127871   .0337575    -0.38   0.705    -.0789672     .053393

...
...

log close
```

If you run the regression with all the control variables the p-value of the F-statistic of the full regression is still insignificant.