. cap cd "C:\Users\Jinhyun\Documents\GitHub\stata\Financial econometrics\Week 1"

.

.

.

.

.

. //open data & install auxiliary package for making tables

. use w1\_data\_experiment1.dta, clear

. //ssc install estout

.

.

.

.

. //2

. desc

Contains data from w1\_data\_experiment1.dta

Observations: 2,188

Variables: 11 26 Sep 2014 11:20

--------------------------------------------------------------------------------------------------

Variable Storage Display Value

name type format label Variable label

--------------------------------------------------------------------------------------------------

treatment byte %8.0g groep info treatment

age float %9.0g Age

ethnminor byte %10.0g allochtoon

Ethnic Minority

ses byte %9.0g socmil Social Economic Status variable measured on 1 - 5

scale

studydur byte %8.0g Study duration (months)

ra byte %8.0g Subjective risk attitude measured on 1 - 10 scale

female byte %9.0g female Female

actrack byte %11.0g wo Academic HE

dr float %9.0g Subjective discount rate per year on [0,1] range

loanexp byte %9.0g loanexp Has prior loan experience dummy

xborrow byte %9.0g Has borrowed after the treatment dummy

--------------------------------------------------------------------------------------------------

Sorted by:

. // xborrow is the outcome variable

. // treat is 0,1 treatment indicator

. // other variables are background characteristics

.

.

.

. //3

. sum

Variable | Obs Mean Std. dev. Min Max

-------------+---------------------------------------------------------

treatment | 2,188 .5018282 .500111 0 1

age | 2,188 21.05757 1.767352 11.47397 34.39178

ethnminor | 2,188 .0447898 .2068893 0 1

ses | 2,188 2.522395 1.384793 1 5

studydur | 2,188 33.01508 13.44552 1 99

-------------+---------------------------------------------------------

ra | 2,188 5.659049 2.073154 1 10

female | 2,188 .6572212 .4747468 0 1

actrack | 2,188 .6096892 .4879314 0 1

dr | 2,188 .2100777 .1857156 .05 .6

loanexp | 2,188 .297989 .4574792 0 1

-------------+---------------------------------------------------------

xborrow | 2,188 .2627971 .4402537 0 1

. drop if age <17

(2 observations deleted)

. // we see that there are some 11 year olds in our data (must be measurement error)

. // assuming they are random coding errors we may drop these observations without

. // introducing a bias.

.

.

.

.

.

.

. //4

. reg treat age ethnminor ses studydur ra female actrack dr loanexp

Source | SS df MS Number of obs = 2,186

-------------+---------------------------------- F(9, 2176) = 0.88

Model | 1.97321475 9 .219246084 Prob > F = 0.5459

Residual | 544.519466 2,176 .250238725 R-squared = 0.0036

-------------+---------------------------------- Adj R-squared = -0.0005

Total | 546.492681 2,185 .250111067 Root MSE = .50024

------------------------------------------------------------------------------

treatment | Coefficient Std. err. t P>|t| [95% conf. interval]

-------------+----------------------------------------------------------------

age | -.0115226 .0071966 -1.60 0.109 -.0256357 .0025904

ethnminor | -.016991 .0524234 -0.32 0.746 -.1197961 .0858141

ses | -.0017121 .007822 -0.22 0.827 -.0170515 .0136272

studydur | .0023554 .000955 2.47 0.014 .0004825 .0042282

ra | -.0006487 .0052403 -0.12 0.901 -.0109252 .0096278

female | -.0211162 .0227609 -0.93 0.354 -.0657516 .0235191

actrack | -.0017736 .0234269 -0.08 0.940 -.047715 .0441679

dr | .003093 .0592451 0.05 0.958 -.11309 .1192759

loanexp | -.0037548 .0239425 -0.16 0.875 -.0507073 .0431976

\_cons | .6909792 .146541 4.72 0.000 .4036043 .9783541

------------------------------------------------------------------------------

. test age=ethnminor=ses=studydur=ra=female=actrack=dr=loanexp=0

( 1) age - ethnminor = 0

( 2) age - ses = 0

( 3) age - studydur = 0

( 4) age - ra = 0

( 5) age - female = 0

( 6) age - actrack = 0

( 7) age - dr = 0

( 8) age - loanexp = 0

( 9) age = 0

F( 9, 2176) = 0.88

Prob > F = 0.5459

. // F-test is not rejected; we have no statistical indication that assignment is

. // conditional on any of these background characteristics, which supports - but

. // does not prove - the claim that assignment was random (p478[526])

.

.

.

.

. //4

. reg xborrow treat

Source | SS df MS Number of obs = 2,186

-------------+---------------------------------- F(1, 2184) = 0.01

Model | .001650327 1 .001650327 Prob > F = 0.9265

Residual | 423.277398 2,184 .193808332 R-squared = 0.0000

-------------+---------------------------------- Adj R-squared = -0.0005

Total | 423.279048 2,185 .193720388 Root MSE = .44024

------------------------------------------------------------------------------

xborrow | Coefficient Std. err. t P>|t| [95% conf. interval]

-------------+----------------------------------------------------------------

treatment | .0017378 .0188319 0.09 0.926 -.0351925 .0386681

\_cons | .261708 .0133405 19.62 0.000 .2355466 .2878694

------------------------------------------------------------------------------

. // the point estimate is near 0 and not significant which mean that the treatment

. // does not significantly increase borrowing. Hence, no indication that a campaign

. // would be usefull.

.

.

. //5

. reg xborrow treat, robust

Linear regression Number of obs = 2,186

F(1, 2184) = 0.01

Prob > F = 0.9265

R-squared = 0.0000

Root MSE = .44024

------------------------------------------------------------------------------

| Robust

xborrow | Coefficient std. err. t P>|t| [95% conf. interval]

-------------+----------------------------------------------------------------

treatment | .0017378 .0188317 0.09 0.926 -.0351922 .0386678

\_cons | .261708 .0133262 19.64 0.000 .2355746 .2878414

------------------------------------------------------------------------------

. est store ols1

. // the standard error decreases a tiny bit. Otherwise nothing is affected;

. // In general standard errors under homoscedasticity will be biased (can be positive

. // or negative) if in fact the variance of the outcome y is related to X, which

. // will be the case in most empirical settings. Therefore we always use

. // heteroscedasticity robust standard errors (see section 5.4 in S&W). This is

. // also the reason that we never calculate the homoscedasticity only F -statistic

. // from p.256[267], but use Stata to compute the appropriate statistic under

. // heteroscedasticity that is given in section 18.3 (this section is not part of the course)

.

. //6

. reg xborrow treat age ethnminor ses studydur ra female actrack dr loanexp, robust

Linear regression Number of obs = 2,186

F(10, 2175) = 56.43

Prob > F = 0.0000

R-squared = 0.2360

Root MSE = .3856

------------------------------------------------------------------------------

| Robust

xborrow | Coefficient std. err. t P>|t| [95% conf. interval]

-------------+----------------------------------------------------------------

treatment | .0037585 .0164897 0.23 0.820 -.0285787 .0360957

age | .0280326 .0069509 4.03 0.000 .0144015 .0416637

ethnminor | .0920709 .0468404 1.97 0.049 .0002144 .1839275

ses | .0024424 .0061818 0.40 0.693 -.0096805 .0145653

studydur | .000286 .0008221 0.35 0.728 -.0013262 .0018982

ra | .0222552 .0039767 5.60 0.000 .0144567 .0300537

female | .0187284 .0174034 1.08 0.282 -.0154005 .0528574

actrack | .0770486 .0178987 4.30 0.000 .0419483 .1121488

dr | .1959725 .048486 4.04 0.000 .1008888 .2910562

loanexp | .3854039 .0215002 17.93 0.000 .3432409 .427567

\_cons | -.6907789 .1360178 -5.08 0.000 -.9575173 -.4240405

------------------------------------------------------------------------------

. est store ols2

. // the estimated treatment effect increases somewhat, but is still very close to 0.

. // the standard error on the estimated treatment effect is a little smaller

. // (0.019->0.016) due to the inclusion of the control variables that absorb part

. // of the unexplained variance. This can also be seen by looking at the R-squared

. // that increases from 0 to 0.23. Hence, including control variables increases

. // precision somewhat, as expected. The included control variables do not have a

. // causal interpretation even though some are significant (see p.233[514])

.

.

.

. //7

. esttab ols1 ols2, keep(treatment) se b(a2)

--------------------------------------------

(1) (2)

xborrow xborrow

--------------------------------------------

treatment 0.0017 0.0038

(0.019) (0.016)

--------------------------------------------

N 2186 2186

--------------------------------------------

Standard errors in parentheses

\* p<0.05, \*\* p<0.01, \*\*\* p<0.001

. // The confidence interval of ols1 is [-.0351922, .0386678], which includes

. // .00375847 similarly the confidence interval of ols2 is [-.0285787, .0360957],

. // which includes .0017378

.

.

.

. //8

. gen begin=(studydur<=24)

. reg xborrow treat c.begin#c.treat begin age ethnminor ses studydur ra female actrack dr loanexp

> , robust

Linear regression Number of obs = 2,186

F(12, 2173) = 47.31

Prob > F = 0.0000

R-squared = 0.2362

Root MSE = .38573

-------------------------------------------------------------------------------------

| Robust

xborrow | Coefficient std. err. t P>|t| [95% conf. interval]

--------------------+----------------------------------------------------------------

treatment | .0147123 .0225894 0.65 0.515 -.0295867 .0590114

|

c.begin#c.treatment | -.0246826 .0331355 -0.74 0.456 -.0896633 .0402981

|

begin | .0114076 .0306932 0.37 0.710 -.0487835 .0715987

age | .0280215 .0069545 4.03 0.000 .0143835 .0416596

ethnminor | .0924925 .0468259 1.98 0.048 .0006643 .1843207

ses | .002572 .0061801 0.42 0.677 -.0095476 .0146916

studydur | .000258 .0011068 0.23 0.816 -.0019124 .0024284

ra | .0222402 .0039787 5.59 0.000 .0144378 .0300425

female | .0188083 .0174195 1.08 0.280 -.0153524 .052969

actrack | .0766522 .0179591 4.27 0.000 .0414334 .1118709

dr | .1950319 .0485335 4.02 0.000 .0998551 .2902088

loanexp | .384946 .0215446 17.87 0.000 .3426958 .4271963

\_cons | -.6945203 .1437986 -4.83 0.000 -.9765175 -.4125231

-------------------------------------------------------------------------------------

. // The coefficient on the interaction term, with a p-value =0.456 , is not significant

. // hence, there is no indication that students at the beginning of their study

. // react more to the treatment than more senior students do.

. // The baseline begin should also be included because it is likely that the

. // begin=0 and begin=1 groups differ in their loan take-up. Not controlling for

. // this will lead to OVB because while treat is randomly assigned, the variable

. // begin\*treat is not. Conditional on begin begin\*treat is randomly assigned, so

. // when we include begin we can also give the coefficient on the interaction

. // begin\*treat a causal interpretation. Of course we can not give the variable begin

. // a causal interpretation because it is not randomly assigned.

. // The c. operator is to tell stata to interpret the variable as continuous.

. // If you were to specify begin#treat stata would automatically interpret the

. // variables as catecorical and include all their interactions.

.

.

. //PIC

. bys begin treat: egen mborrow = mean(xborrow)

. twoway (scatter xborrow treatment) ///

> (scatter mborrow treatment if begin==0, ms(O) msize(large) mc(green)) ///

> (scatter mborrow treatment if begin==1, ms(x) msize(huge) mc(blue)) ///

> (lfit xborrow treatment if begin==0, lc(green)) ///

> (lfit xborrow treatment if begin==1, lc(blue)), ///

> legend(label(1 "raw data") label(4 "begin==0") ///

> label(5 "begin==1") order(1 4 5) col(3)) ytitle(xborrow) name(loaninfo, replace)

.

.

.

.

.

. //9

. use w1\_data\_experiment2.dta, clear

. desc

Contains data from w1\_data\_experiment2.dta

Observations: 800

Variables: 6 26 Sep 2014 11:21

--------------------------------------------------------------------------------------------------

Variable Storage Display Value

name type format label Variable label

--------------------------------------------------------------------------------------------------

studentid double %10.0g Individual student ID

major double %10.0g Student follows econometrics course as major

male double %10.0g

homework double %10.0g Student has recieved compusory homework treatment

examscore double %10.0g Grade obtained at econometrics exam

dropout double %10.0g Student drops out at end of year

--------------------------------------------------------------------------------------------------

Sorted by:

. // Examscore and dropout are outcome variables that are realized after the treatment

. // homework is 0,1 treatment indicator studentid id student identifier other variables

. // are background characteristics

.

.

. //10

. sum

Variable | Obs Mean Std. dev. Min Max

-------------+---------------------------------------------------------

studentid | 800 4.86e+07 2.95e+07 160369 1.00e+08

major | 800 .695 .4606955 0 1

male | 800 .57375 .4948404 0 1

homework | 800 .4875 .5001564 0 1

examscore | 800 6.9225 1.163265 3.5 10

-------------+---------------------------------------------------------

dropout | 800 .30375 .4601637 0 1

. // All variables seem within range, no indication of outliers

.

. //11

. reg examscore homework

Source | SS df MS Number of obs = 800

-------------+---------------------------------- F(1, 798) = 62.44

Model | 78.4555378 1 78.4555378 Prob > F = 0.0000

Residual | 1002.73946 798 1.25656574 R-squared = 0.0726

-------------+---------------------------------- Adj R-squared = 0.0714

Total | 1081.195 799 1.35318523 Root MSE = 1.121

------------------------------------------------------------------------------

examscore | Coefficient Std. err. t P>|t| [95% conf. interval]

-------------+----------------------------------------------------------------

homework | .6265166 .0792891 7.90 0.000 .4708768 .7821564

\_cons | 6.617073 .0553606 119.53 0.000 6.508404 6.725743

------------------------------------------------------------------------------

. est store ols3

. // It seems that the compulsory homework assignment increases the exam score by

. // about 0.62 on a scale from 0-10 which is quite substantial. In terms of standard

. // deviations the effect is 0.62/1.16=.53, which is a verly large effect.

.

.

.

. //12

. tab homework major, col nofreq

Student |

has |

recieved | Student follows

compusory | econometrics course

homework | as major

treatment | 0 1 | Total

-----------+----------------------+----------

0 | 62.30 46.40 | 51.25

1 | 37.70 53.60 | 48.75

-----------+----------------------+----------

Total | 100.00 100.00 | 100.00

. // The table shows that non majors have about 38% change of being treated, while

. // majors recieve treatment in 53% of cases; This clearly indicates that the

. // homework treatment is selective with respect to major status

. reg homework major male

Source | SS df MS Number of obs = 800

-------------+---------------------------------- F(2, 797) = 9.72

Model | 4.75816105 2 2.37908052 Prob > F = 0.0001

Residual | 195.116839 797 .244814102 R-squared = 0.0238

-------------+---------------------------------- Adj R-squared = 0.0214

Total | 199.875 799 .250156446 Root MSE = .49479

------------------------------------------------------------------------------

homework | Coefficient Std. err. t P>|t| [95% conf. interval]

-------------+----------------------------------------------------------------

major | .1686755 .0386349 4.37 0.000 .0928373 .2445138

male | .0501135 .035969 1.39 0.164 -.0204918 .1207187

\_cons | .3415179 .040666 8.40 0.000 .2616929 .421343

------------------------------------------------------------------------------

. // The F-statistic of the regression of covariates on treatment status is 9.7, with

. // pvalue=0.0001; hence we reject the H0 that there is no relation between background

. // covariates in favour of the alternative hypothesis that treatment status is selective

. // with respect to one or more of the covariates.

. // Judging from the table it is only the variable major that is related to treatment,

. // the male variable is not. Both can be included in the regression. The major variable

. // HAS TO BE INCLUDED **as a control variable** in the regression to get unbiased estimates of the causal

. // effect of homework on grades. The male variable can be included in increase

. // **statistical precision**

.

.

. //13

. reg examscore homework major male

Source | SS df MS Number of obs = 800

-------------+---------------------------------- F(3, 796) = 78.35

Model | 246.491305 3 82.1637683 Prob > F = 0.0000

Residual | 834.703695 796 1.04862273 R-squared = 0.2280

-------------+---------------------------------- Adj R-squared = 0.2251

Total | 1081.195 799 1.35318523 Root MSE = 1.024

------------------------------------------------------------------------------

examscore | Coefficient Std. err. t P>|t| [95% conf. interval]

-------------+----------------------------------------------------------------

homework | .4937244 .0733098 6.73 0.000 .3498209 .6376278

major | .9934783 .0809103 12.28 0.000 .8346556 1.152301

male | -.0548427 .074533 -0.74 0.462 -.2011471 .0914617

\_cons | 6.022808 .0878082 68.59 0.000 5.850445 6.195171

------------------------------------------------------------------------------

. est store ols4

. // The coeff on homework has dropped from .63 to .49 (i.e. by about 22%) and now

. // has a causal interpretation. Even though major status is a significant and strong

. // predictor of grades, we can not give it a causal interpretation because we have

. // not randomly assigned students to following the course as part of their major

. // or just an elective. The variable male also has no causal interpretion (and

. // it is actually hard to think of a situation where this could be randomized) and

. // it is not significant. Therefore it will not contribute much to increasing precision

. esttab ols3 ols4, se b(a2)

--------------------------------------------

(1) (2)

examscore examscore

--------------------------------------------

homework 0.63\*\*\* 0.49\*\*\*

(0.079) (0.073)

major 0.99\*\*\*

(0.081)

male -0.055

(0.075)

\_cons 6.62\*\*\* 6.02\*\*\*

(0.055) (0.088)

--------------------------------------------

N 800 800

--------------------------------------------

Standard errors in parentheses

\* p<0.05, \*\* p<0.01, \*\*\* p<0.001

.

.

. //14

. // The confidence interval of ols3 is [.4708768, .7821564], which includes .4937244

. // similarly the confidence interval of ols4 is [.3498209, .6376278], which

. // includes .6265166. Hence, this rough eyeballing test does not seem to indicate

. // both estimates are siginficantly different

.

.

. //15

. qui suest ols3 ols4

. lincom [ols4\_mean]homework-[ols3\_mean]homework

( 1) - [ols3\_mean]homework + [ols4\_mean]homework = 0

------------------------------------------------------------------------------

| Coefficient Std. err. z P>|z| [95% conf. interval]

-------------+----------------------------------------------------------------

(1) | -.1327922 .0337088 -3.94 0.000 -.1988602 -.0667242

------------------------------------------------------------------------------

. // The output shows that the coefficient has dropped with -0.13 and significantly

. // so. The reason we do reject equality of the two estimates now, is that the

. // covariance between the two estimators is positive (you can calculate that it

. // is equal .0052 using the formula), which REDUCES the standard error of the

. // difference. Hence, the standard error of the difference is of the magnitude of

. // 0.03, much smaller than the s.e.'s of the individual coefficients (about 0.08).

. di 0.5\*(0.079^2 + 0.073^2 - .0337^2)

.00521715

.

. // The implication is that the first estimate was upward biased. This can be

. // explained by the fact that majors are probably more able and motivated doing

. // this course than the electives, i.e. they would have performed better anyway.

. // Because the treatment group contains more of them, we overstimate the effect

. // on grades if we do not correct for this (i.e. if we do not include it as a

. // control variable)

.

.

.

. //15

. reg examscore homework major male dropout

Source | SS df MS Number of obs = 800

-------------+---------------------------------- F(4, 795) = 66.87

Model | 272.184286 4 68.0460715 Prob > F = 0.0000

Residual | 809.010714 795 1.01762354 R-squared = 0.2517

-------------+---------------------------------- Adj R-squared = 0.2480

Total | 1081.195 799 1.35318523 Root MSE = 1.0088

------------------------------------------------------------------------------

examscore | Coefficient Std. err. t P>|t| [95% conf. interval]

-------------+----------------------------------------------------------------

homework | .4202277 .0736845 5.70 0.000 .2755886 .5648669

major | 1.027045 .0799848 12.84 0.000 .8700383 1.184051

male | -.0403041 .07348 -0.55 0.583 -.1845419 .1039337

dropout | -.3982276 .0792534 -5.02 0.000 -.5537982 -.242657

\_cons | 6.147929 .0900134 68.30 0.000 5.971237 6.324621

------------------------------------------------------------------------------

. // Including the variable dropout leads to a further drop in the coefficient.

. // This result cannot be given a causal interpretation, however, because dropout

. // is also an OUTCOME variable. Inclusion of another outcome variable re-introduces

. // endogeneity when (1) the other outcome is affected by the treatment (meaning

. // it is truely an outcome that is being affected) and (2) There are unobserved

. // factors that influence both outcomes.

. //

. // In this particular case it seems likely that unobserved ability/motivation

. // will both affect the dropout decision and the examgrade. When the dropout

. // decision is also affected by homework, inculding it into the regression on

. // grades will introduce endogeneity EVEN THOUGH homework is randomly assigned.

. // This is because homework and the dropout decision are related (homework -> dropout),

. // and the dropout decision is related to unobserved ability (unobserved ability

. // -> dropout). Hence, this will create a relation between homework and unobserved

. // ability, which is not what we want (homework -> dropout <- unobserved ability).

. // [the true argumentation is more technical, see Canvas Q3, but this is the way

. // you can think about it intuitively]

.

.

end of do-file

.