RCT Implementation

Randomized Control Trial

In this excecise, I'll provide some simple examples of how to implement a basic RCT analysis.

Before we start, however, we need to create some data.

1. Estimate two wages models. For men and women, using an heteroskedastic regression model.

The goal: estimate two potential outcomes for wages. One following Womens wage structure, and the other using Men wage structure.

```
set linesize 255
frause oaxaca, clear
set seed 101
qui:hetreg lnwage age agesq married divorced kids6 kids714 if female==0, het(age agesq mar
est sto m1
predict xb1, xb
predict s1, sigma
qui:hetreg lnwage age agesq married divorced kids6 kids714 if female==1, het(age agesq mar
est sto m2
predict xb0, xb
predict s0, sigma
```

```
(Excerpt from the Swiss Labor Market Survey 1998)
```

Here, as I did for UQregression, I obtain the predicted mean and predicted standard errors. and simulate two wages. I will use wage differences between simulated men and women predicted log wages as the treated effect

$$TE = \hat{lnw}_1 - \hat{lnw}_0$$
$$\hat{lnw}_1 \sim N(xb_1, s_1^2); \hat{lnw}_0 \sim N(xb_0, s_0^2)$$

```
gen lnwage1 = rnormal(xb0,s0)
gen lnwage0 = rnormal(xb1,s1)
gen teff=lnwage1-lnwage0
```

I will create a randomzed treatment, and assume those treated get the treatment effect

$$Y_i = Y_i(0) + trt_i * \delta_i$$

```
gen trt = runiform()<.5
replace lnwage = lnwage0 + trt * teff
```

```
(1,647 real changes made)
```

Visual Exploration

Now that we have a randomized treatment, we could start exploring the data:

```
two (kdensity lnwage if trt == 1) (kdensity lnwage if trt == 0) , ///
    legend(order(1 "Treated" 2 "Untreated"))
```

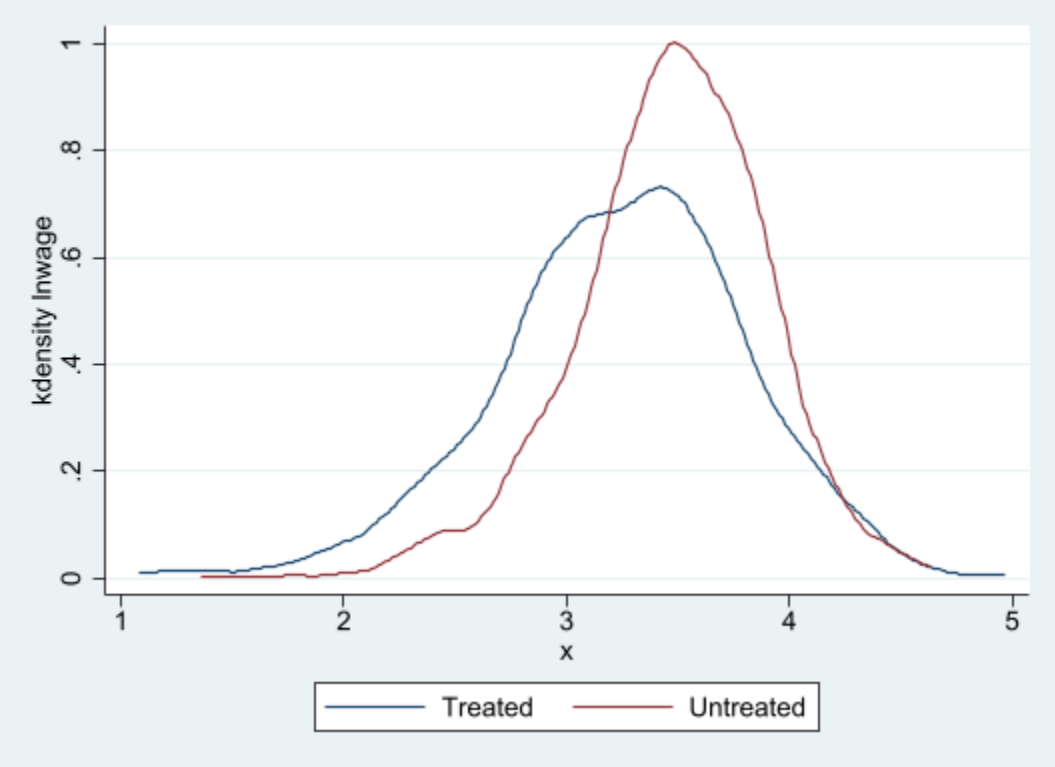

Figure 1: Log wage distribution between Treated and untreated

In order to estimate the treatment effects, we could simple estimate a regression model of the outcome. Compare it to the treatment effect

Estimation of ATE Effect

```
** True Effect
sum teff
** Simple Regression
set linesize 255
reg lnwage  trt, robust
est sto m0
```

```
    Variable |        Obs        Mean    Std. dev.       Min        Max
-------------+---------------------------------------------------------
        teff |      1,647   -.2121816    .6613419  -3.024343   2.704082
```

```
Linear regression                               Number of obs     =      1,647
                                                F(1, 1645)        =      79.85
                                                Prob > F          =     0.0000
                                                R-squared         =     0.0464
                                                Root MSE          =      .5017

------------------------------------------------------------------------------
             |               Robust
      lnwage | Coefficient  std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
         trt |
  -.2213292   .0247679    -8.94   0.000    -.2699092   -.1727492
       _cons |   3.465982   .0150123   230.88   0.000     3.436537    3.495427
------------------------------------------------------------------------------
```

Because treatment is randomized, we could also add other controls to the model, and improve on precision

```
qui:reg lnwage  trt age agesq , robust
est sto m1
qui:reg lnwage  trt age agesq married divorced , robust
est sto m2
qui:reg lnwage  trt age agesq married divorced kids6 kids714 , robust
est sto m3

esttab m0 m1 m2 m3, se nonum mtitle("m0" "m1" "m2" "m3") keep(trt) md
```

	m0	m1	m2	m3
trt	-0.221***	-0.214***	-0.213***	-0.212***
	(0.0248)	(0.0225)	(0.0225)	(0.0224)
N	1647	1647	1647	1647

Standard errors in parentheses * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Falsification

We could just use other outcomes that shouldnt be affected by the treatment. You expect they have no impact on outcome

```
qui:reg exper  trt age agesq married divorced kids6 kids714 , robust
est sto m0
qui:reg tenure trt age agesq married divorced kids6 kids714 , robust
est sto m1
esttab m0 m1 , se nonum mtitle("m0" "m1") keep(trt) md
```

	m0	m1
trt	-0.104	-0.458
	(0.366)	(0.335)
N	1434	1434

Standard errors in parentheses * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Balance test

You should also try to create balance tables, where you compare and test if characteristics are similar across treated and control groups:

```
tabstat age agesq married divorced kids6 kids714 , by(trt)
sureg age agesq married divorced kids6 kids714 =trt,
```

```
Summary statistics: Mean
Group variable: trt

     trt |       age     agesq   married  divorced     kids6   kids714
---------+--------------------------------------------------------------
       0 |  39.14475   1649.63    .53076  .1206273  .2979493  .3365501
       1 |   39.3643  1675.521  .5158924  .1466993  .2713936  .3215159
---------+--------------------------------------------------------------
   Total |  39.25379  1662.489  .5233758  .1335762  .2847602  .3290832
------------------------------------------------------------------------
```

```
Seemingly unrelated regression
------------------------------------------------------------------------------
Equation             Obs   Params         RMSE  "R-squared"      chi2   P>chi2
------------------------------------------------------------------------------
age                1,647        1     11.02798      0.0001       0.16   0.6862
agesq              1,647        1     893.7224      0.0002       0.35   0.5566
married            1,647        1     .4993979      0.0002       0.36   0.5458
divorced           1,647        1     .3399466      0.0015       2.42   0.1197
kids6              1,647        1     .6626276      0.0004       0.66   0.4161
kids714            1,647        1     .7071256      0.0001       0.19   0.6662
------------------------------------------------------------------------------

------------------------------------------------------------------------------
             | Coefficient  Std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
age          |
         trt |   .2195505   .5434865     0.40   0.686    -.8456636    1.284765
       _cons |   39.14475   .3830175   102.20   0.000     38.39405    39.89545
-------------+----------------------------------------------------------------
agesq        |
         trt |   25.89111   44.04489     0.59   0.557    -60.43528    112.2175
       _cons |    1649.63   31.04026    53.14   0.000     1588.792    1710.467
-------------+----------------------------------------------------------------
married      |
         trt |  -.0148675   .0246116    -0.60   0.546    -.0631054    .0333703
       _cons |     .53076   .0173448    30.60   0.000     .4967648    .5647552
-------------+----------------------------------------------------------------
divorced     |
         trt |    .026072   .0167534     1.56   0.120    -.0067641    .0589081
       _cons |   .1206273   .0118068    10.22   0.000     .0974863    .1437682
-------------+----------------------------------------------------------------
kids6        |
         trt |  -.0265557    .032656    -0.81   0.416    -.0905602    .0374488
       _cons |   .2979493    .023014    12.95   0.000     .2528427     .343056
-------------+----------------------------------------------------------------
kids714      |
         trt |  -.0150342   .0348489    -0.43   0.666    -.0833368    .0532685
       _cons |   .3365501   .0245595    13.70   0.000     .2884143    .3846858
------------------------------------------------------------------------------
```

Here, the goal is just to see if `trt` is not-significant across groups