THE UNIVERSITY OF CHICAGO
Harris Public Policy

Harris School of Public Policy
PPHA 346 Program Evaluation
Winter 2018

Assignment 2

Due date: Feb. 6, 2018, by 11:59pm via Canvas. Your answers must be typed; submit a .doc or .pdf document (.html is fine if you use Python).

Format: please submit your answers in the following format. You may copy-paste Stata outputs by simply select, right click, copy as picture.

For example:

1. Question

Written explanation

`stata code`

Stata output

1. What is the mean of X?

The mean of X is 5.33.

```
. mean X

Mean estimation                   Number of obs   =     21646

--------------------------------------------------------------
             |       Mean   Std. Err.     [95% Conf. Interval]
-------------+------------------------------------------------
           X |   5.331285    .0067868      5.317983    5.344588
--------------------------------------------------------------
```

This will make you show every line of your code throughout your answers, so you do not have to also submit a do-file. Some questions will only need code and no outputs, so in those cases simply present the code you typed.

Non-Stata Languages:
You are welcome to use R, Python, SAS, etc., rather than Stata if that's your preference. Just keep in mind that if you do:

- Answer keys and do-files for answer keys will be in Stata, so if you get a wrong answer they probably won't be helpful to know what you did wrong.
- Your grade will be based entirely on whether you got the right numeric answer and whether your written explanations/interpretations were correct. Stata users who get a wrong numeric answer due to a minor mistake in their code might still get partial credit if their reasoning and interpretation were right, but we might not be able to identify glitches in code for non-Stata users.
- Regardless of the language you use, you are always required to show your code after your written explanations and before your outputs, as indicated above.

Late Submissions:
Please turn in your assignment on time. Per syllabus policy, if you submit after the deadline, two percentage points of your grade (10 percentage points max) will be discounted. This policy did not apply for assignment 1, but it will apply for assignments 2, 3, and 4.

Does Your Neighborhood Matter?

Research has long suggested that living in disadvantaged neighborhoods has a negative effect on people's earnings, education, health, crime involvement, and other outcomes. Since there are several selection problems around who lives where, a randomized experiment is a good tool to isolate the causal effects that living in a given neighborhood has on people. During 1994-1998, the US Department of Housing and Urban Development enrolled families in the Moving to Opportunity (MTO) program, which offered housing vouchers via random lottery to families with children living in high-poverty areas so that they could move to less-disadvantaged neighborhoods. The program was implemented in Baltimore, Boston, Chicago, Los Angeles, and New York City. Families were randomized into three groups:

i. *Experimental group*. The vouchers for this group subsidized private-market rents, but families could only use them to move to neighborhoods where the poverty rate was below 10%, in an effort to assure they moved to significantly better neighborhoods.
ii. *Section 8 group*. These vouchers also subsidized rent, but there were no restrictions on what kind of neighborhood the families could move to.
iii. *Control group*. These families were not offered any vouchers.

A group of evaluators found that 10-15 years after implementing the program, MTO improved some physical and mental health outcomes of the adults living in the household, but it did not have an effect on their economic self-sufficiency, among other findings. You will work with a subset of these evaluators' individual-level data.[1]

Open `MTO_data.dta` and type `describe` in Stata to read the description of each variable.

The post-treatment outcome variables are:

```
overall_index_z
economic_self_sufficiency_z
absence_phys_probs_z
absence_mental_probs_z
psychol_distress_z
```

The first is an index encompassing several variables for overall physical and mental health (lower values are worst health, higher values better). The second is an index for how economically sufficient the individual is (lower values are less sufficient and higher values are more sufficient). The last three indicate the absence of physical health problems, absence of mental health problems, and the presence of psychological distress. Note that these variables end with `_z`, which means that they have been conveniently transformed to *Z*-scores so that you can easily interpret the changes observed in either the experimental or the section 8 groups in terms of standard deviations away from the control group. Hence, if you obtain that the treatment given to the experimental group has an impact of, for example, $\hat{\beta}_1 = -1.2$ on `psychol_distress_z`, it suggests that people with this treatment had a mean psychological distress index that was 1.2 standard deviations below the control group's mean index.

---

[1] If you want more background information, you can find the evaluation here: http://www.nber.org/mtopublic/final/MTO_AERPP_2013.pdf. However, you are not required to read the evaluation; these instructions and the dataset are meant to be self-contained. Because you will be working with an altered subset of the evaluators' data, you will not get the same results that they do. Do not expect to find similar estimates if you decide to read it.

The variables indicating which group individuals were randomized into are:

```
group_experimental
group_section8
group_control
```

These are simple dummies equal to 1 if the individual belongs to that group, 0 if otherwise.

The variables from `city_baltimore` to `applied_sec8_before` are (pre-treatment) baseline characteristics. For now, ignore the variable `complier`.

Part I

1. Perform a balance test on pre-treatment characteristics by completing the table below, rounding to three decimals. You may perform simple t-tests (`ttest`); use the default equal-variances assumption.

	(i)	(ii)	(iii)	(iv)	(v)	(vi)	(vii)
	Experimental group (E) mean	Section 8 group (S) mean	Control group (C) mean	Difference E - C	*p*-value E - C	Difference S - C	*p*-value S - C
`male`	.495						
`head_earnings`							.729
`bmi_40`				-.020			
`no_family_in_nbhd`			.634				
`anxiety`		.353					

2. Based on the table above only, what can you say about the program's random lottery that was used to assign vouchers?

3. Show the same results as columns (iv)-(vii) from the table by using bivariate regressions.

4. What share of the observations corresponds to the experimental, the section 8, and the control groups?

5. Calculate the average treatment effect for the experimental group manually by comparing two means, for each of the five post-treatment outcome variables. Round to three decimals. Interpret each result and complete the following table.

Experimental Group		
Outcome	**ATE**	**Interpretation**
`overall_index_z`		
`economic_self_sufficiency_z`		
`absence_phys_probs_z`	.066	
`absence_mental_probs_z`		
`psychol_distress_z`		

6. Calculate the average treatment effect for the section 8 group manually by comparing two means, for each of the five post-treatment outcome variables. Round to three decimals and complete the following table. Do you get a difference in the direction of any effect compared to the experimental group's effects? Provide at least one possible idea for why this might be the case.

Section 8 Group	
Outcome	**ATE**
`overall_index_z`	
`economic_self_sufficiency_z`	
`absence_phys_probs_z`	
`absence_mental_probs_z`	
`psychol_distress_z`	

7. Replicate the same ATEs from question 5, now using bivariate regressions. What do you learn about the precision of your estimates? Which are significant at the 95% level? Which at the 90% level?

8. Replicate the same ATEs from question 6, now using regressions. What can you say about section 8's estimates' precision relative to the experimental group? Which are significant at what level?

9. Given the R^2 and the precision of your estimates in questions 7 and 8, is there something you would do differently?

10. For simplicity, add the entire set of pre-treatment covariates from `city_baltimore` to `applied_sec8_before` as controls on your regression for the experimental group. Note that some of the dummies are mutually exclusive and create multicollinearity, so do not include the following ones: `city_NYC, age_1_35, no_high_school,` and also `diabetes`. The command `global` might save you time, but you're not required to use it. Report only your code for this question, not your entire output, but complete the following tables with information from your output.

EXPERIMENTAL GROUP					
	Overall index (z)	Economic self-sufficiency (z)	Absence physical problems (z)	Absence mental health problems (z)	Psychological distress (z)
`group_experimental` $\hat{\delta}$				0.0301***	
s.e.					
…	…	…	…	…	…
$N =$					
$R^2 =$					
Adj. $R^2 =$			0.966		

$\hat{\delta}$ is the slope coefficient for the average treatment effect.
s.e. is the standard error of the slope coefficient.

11. Add the same covariates now to your regressions for the section 8 group. Complete the table below.

SECTION 8 GROUP					
	Overall index (z)	Economic self-sufficiency (z)	Absence physical problems (z)	Absence mental health problems (z)	Psychological distress (z)
`group_section8` $\hat{\delta}$					
s.e.				(0.0166)	
…	…	…	…	…	…
N =					
R^2 =					
Adj. R^2 =	0.961				

$\hat{\delta}$ is the slope coefficient for the average treatment effect.
s.e. is the standard error of the slope coefficient.

12. How much do your estimates change from question 7 to 10 (experimental) and from 8 to 11 (section 8)? How did their standard errors change? Recall slides 9-10 from lecture 3 (Social Experiments I) and explain why you see or do not see changes on the estimates and their standard errors. Is there anything that you would have done in addition to what you did for question 1? How does this relate to why your estimates changed or did not change?

13. How do you call this approach to estimating treatment effects given that you ignored the variable `complier`? Describe the treatment that you were really analyzing here.

Part II

Now, consider the variable `complier` in your analysis (and ignore the long list of controls for this entire Part II). `Complier` is a dummy variable that equals 1 if the individual is a complier and 0 if she is not, which means something different depending on what group the individual belongs to. If you were randomized into either the experimental group or the section 8 group and you are also a complier, it means that you were not only offered the voucher, but you also actually moved to a different neighborhood. However, if you were randomized into the control group and you are also a complier, it means that you complied with your random assignment and indeed did not move to a different neighborhood.

14. What share of the experimental group actually moved? How about the section 8 and the control groups? Calculate the dropout rate in each treatment group and the rate of control substitution in the control group.

15. For both the experimental and the section 8 groups, manually compute the Bloom estimator for the outcome variable `overall_index_z`. How do these compare to the ATEs that you obtained for this outcome in questions 5 and 6? How do you interpret these estimators?

16. Consider this question an introduction to instrumental variables (which you will learn later in the course). You will replicate the same estimators as in question 15 but using regressions.

    a) Generate a dummy variable equal to 1 if the individual belongs to the experimental group and is also a complier, and equal to 0 if otherwise. As a first stage, regress this dummy on the indicator for random assignment into the experimental group. What is this estimator equal to?

    b) Store the predicted values from the linear model in (a) (`predict` is your friend). As a second stage, regress the outcome variable, `overall_index_z`, on the predicted values you stored. What is the estimate you obtain equal to?

    c) Repeat the same steps for the section 8 group. Report your code and estimates.

17. What is the main assumption of the Bloom estimator?

18. Does the numerator in your Bloom estimator remind you of another estimator you've seen before in class? Under what circumstances will that estimator be unbiased?