

FLS 6415: Class 3 Homework

August 29, 2017

Remember to answer all the questions in R markdown and produce a PDF. Email your completed homework (R markdown file and PDF) to jonnyphillips@gmail.com by midnight the night before class.

First, some very quick questions to encourage you to explore the app that simulates potential outcomes and randomized experiments. Use the app at https://poliong.shinyapps.io/causation_app/ to answer the following questions:

1. Set the number of units to at least 20. Try generating random units a few times to understand the patterns. Comparing the ‘Actual’ Average Treatment Effect with the ‘Observed’ ATE, which treatment assignment mechanism appears to provide an unbiased estimate? Which provides an over-estimate, and which an under-estimate?
2. Using the table of potential outcomes (and, if helpful, the ‘data’ tab), explain why the ‘self-selection’ treatment assignment mechanism does not accurately estimate the actual ATE.
3. On the ‘Analysis’ tab, why does the comparison of means estimate exactly the same ATE as the regression?
4. Set a high number of units and random treatment assignment. On the ‘Repeated Experiments’ tab, describe how the bias of the observed ATE compared to the actual ATE changes as the number of repeated experiments increases from 1 to 100.

The questions below focus on replicating the results of the analysis in Olken (2010). While you should try and conduct the same analysis and see if you reach the same conclusion as Olken, it is NOT necessary that you get exactly the same numerical result. Try, but if you can’t that’s okay. Similarly, don’t worry about copying the same formatting of the tables - as long as the results are clear. The aim is to focus on the understanding, skills and coding knowledge needed to apply the same type of analysis.

A dictionary of variable names and description for each table is provided at the end of this document, along with code you can copy and paste into your own analysis to save you time typing.

1. Open the file `clean_table2_village_data.dta`. Implement the balance test that Olken conducts on pages 247-248 to compare the village population in treatment and control villages.
2. Conduct a simpler balance test on village population by using a t-test to assess the difference in means for village population between treatment and control villages. How does the result differ from the regression in Q1? What explains the difference?
3. Reproduce the full balance table (Table 2 in Olken) using the files `clean_table2_village_data.dta` and `clean_table2_hh_data.dta`. (You can make separate tables for the village and respondent characteristics. Try to report the summary tables, but if you can’t just report each regression separately. In your output there’s no need to include the mean of the meeting group, the standard errors in brackets or the number of observations column; just the coefficient and p-value is fine).
4. Interpret the balance table. What does this table tell us about (i) the randomization assignment process, and (ii) the prospects for causal inference?
5. Open the file `clean_table7_10_data.dta`. The following variables are currently scaled so that they can take on integer values between 1 and 4: `sat_prop`, `benefit`, `fair`, `aspire`, `kdp_satw2`. Transform these variables so that they are on a scale between 0 and 1.
6. Open the file `clean_table7_10_data.dta` and get as close as possible to replicating column (1) of Table 7 using OLS. Use the transformed versions of the outcome variables you generated in Q5. The control variables are `phase2`, `male`, `age`, `npers`, `c_servant`, `enterp`, `hwife`, `laborer`, `not-in_if`, `other`, `priv_comp`,

student, teacher, trader, hexpcap. (Note there is an error that means the coefficient values for ‘benefit’ and ‘use’ you calculate will not be the same as in Olken’s table (see Q9)).

7. Replicate column (1) of Table 7 again but without including the control variables (but keep variable phase2). Does this suggest any different conclusions about the average treatment effect compared to the regression with controls? Why?
8. For the variable ‘use’ (“Will you use the project?”), conduct a t-test for the differences in means between the treatment and control groups. Report the difference in means and the p-value of the t-test. Does this suggest any different conclusions about the average treatment effect compared to the regression without controls? Why?
9. (Challenging) Investigate why the coefficient values for ‘benefit’ and ‘use’ in Q7 are different to those in Olken’s Table 7. It may help to refer to the stata do file *100305tables.do*.

Variables

`clean_table2_village_data.dta`

```
library(knitr)
kable(tab2_vill[,c(1,2)])
```

Variable	Description
desa_pop	Village population (1,000 inhabitants)
ag_wage	Agricultural wage (1,000 Rupiah)
asphalt	Percent village roads that are asphalt
dusun_count	Number of hamlets per village
church_mosque	Number of churches and mosques per village
dist_capital	Distance to subdistrict capital (km)
v_ethnic_frag	Village ethnic fragmentation
v_rel_frag	Village religious fragmentation
vh_age	Village head age
educ	Village head years of education
v_numvhcandidates	Number of village head candidates in last village head election
v_morethan1candidate	More than one candidate in last village head election
v_sharevhvoted	Share of population that voted in last village head election
v_vhvictorymargin	Village head’s margin of victory in last election (if challenger)
v_numapararat	Number of village government executive branch members
v_sharedusunsinapararat	Share of hamlets represented in village executive branch
bpd_numpeople	Number of people in village parliament
v_sharedusunsinbpd	Share of hamlets represented in village parliament
bpd_nummeetings	Number of village parliament meetings in last year
bpd_districtsystem	Village parliament district system (1 = district, 0 = at large)
ppk_numproj	Number of previous KDP projects

`clean_table2_hh_data.dta`

```
kable(tab2_indi[,c(1,2)])
```

Variable	Description
zlhhexpcap	Survey respondent predicted log per capita expenditure

Variable	Description
educ	Survey respondent years education
female	Survey respondent is female
age	Survey respondent age
farmer_buruh	Survey respondent is farmer

clean_table7_10_data.dta

```
kable(tab7[,c(1,2)])
```

Variable	Description
sat_prop	Was the project chosen in accordance with your wishes?
benefit	Will the proposal benefit you personally?
use	Will you use the project?
fair	Was the chosen proposal fair?
aspire	Is the chosen proposal in accordance with the people's aspirations?
kdp_satw2	Are you satisfied with the KDP?
jobapproval_pres	Job approval of president of Indonesia
jobapproval_vh	Job approval of village head

```
t2_vars_vill <- c("desa_pop", "ag_wage", "asphalt", "dusun_count", "church_mosque",
  "dist_capital", "v_ethnic_frag", "v_rel_frag", "vh_age", "educ", "v_numvhcandidates",
  "v_morethan1candidate", "v_sharevhvoted", "v_vhvictorymargin", "v_numapararat",
  "v_sharedusunsinapararat", "bpd_numpeople", "v_sharedusunsinbpd", "bpd_nummeetings",
  "bpd_districtsystem", "ppk_numproj")
t2_names_vill <- c("Village population (1,000 inhabitants)", "Agricultural wage (1,000 Rupiah)",
  "Percent village roads that are asphalt", "Number of hamlets per village",
  "Number of churches and mosques per village", "Distance to subdistrict capital (km)",
  "Village ethnic fragmentation", "Village religious fragmentation", "Village head age",
  "Village head years of education", "Number of village head candidates in last village head election",
  "More than one candidate in last village head election", "Share of population that voted in last vi",
  "Village head's margin of victory in last election (if challenger)", "Number of village government c",
  "Share of hamlets represented in village executive branch", "Number of people in village parliament",
  "Share of hamlets represented in village parliament", "Number of village parliament meetings in las",
  "Villiage parliament district system (1 = district, 0 = at large)", "Number of previous KDP project")

tab2_vill <- data.frame(cbind(t2_vars_vill, t2_names_vill))
colnames(tab2_vill) <- c("Variable", "Description")

t2_vars_indi <- c("zlhhexpcap", "educ", "female", "age", "farmer_buruh")
t2_names_indi <- c("Survey respondent predicted log per capita expenditure",
  "Survey respondent years education", "Survey respondent is female", "Survey respondent age",
  "Survey respondent is farmer")
tab2_indi <- data.frame(cbind(t2_vars_indi, t2_names_indi))
colnames(tab2_indi) <- c("Variable", "Description")

t7_vars <- c("sat_prop", "benefit", "use", "fair", "aspire", "kdp_satw2", "jobapproval_pres",
  "jobapproval_vh")
t7_names <- c("Was the project chosen in accordance with your wishes?", "Will the proposal benefit you p",
  "Will you use the project?", "Was the chosen proposal fair?", "Is the chosen proposal in accordance
```

```
"Are you satisfied with the KDP?", "Job approval of president of Indonesia",  
"Job approval of village head")  
tab7 <- data.frame(cbind(t7_vars, t7_names))  
colnames(tab7) <- c("Variable", "Description")
```