Methods III: Replication Exercise 1

April, 2019

The data for Gerber, Green and Larimer (GGL, 2008) is available on the course website. You should first read through GGL 2008 quickly to understand the context of the field experiment.

Download the data from the course website and answer the following questions in either R or Stata. You should submit (i) your code, and (ii) a document or PDF containing your answers to jonnyphillips@gmail.com by midnight on Wednesday 10th April. If you get stuck, please feel free to email me, or in the worst case skip that question and continue to the next.

For the replication, don't worry about copying the specific details or formatting of the tables - as long as the results are clear. If you are using R, I encourage you to use R markdown to make it easy to combine your code and answers.

Here is a table of variable names and descriptions from the dataset:

Variable	Description
hh_id	Household Identifier
hh_size	Household Size
cluster	Block
treatment	Treatment status
g2000	Voted in 2000 General Election
g2002	Voted in 2002 General Election
p2000	Voted in 2000 Primary Election
p2002	Voted in 2002 Primary Election
p2004	Voted in 2004 Primary Election
sex	Sex
age	Age
voted	Voted in 2006 Primary Election

1. What hypotheses are GGL testing? Where did they get these hypotheses from?

The hypotheses are whether civic duty motivations affect voting turnout, and whether the mechanism is intrinsic motivation or social shaming. These hypotheses come from the existing literature and the predictions of a formal model.

2. What are the treatment and control conditions? What is the outcome variable?

Treatment is the receipt of one of four messages in the mail. Control is receiving no message at all. Outcome is whether you voted in the 2006 Primary election.

3. What is the unit of analysis? What is the level at which treatment is applied?

The individual. Treatment was applied at the level of the household.

4. Did randomization work? Let's reproduce Table 1 of GGL to conduct balance tests between the treatment and control groups on pre-treatment covariates. Note that GGL evaluate balance at the *household* level so you first need to aggregate the individual data to the household level by finding the household mean on each of the variables we want to assess balance for. Then calculate the mean across household separately for the control and treatment groups. What do we learn from the results?

```
d <- read_csv("GGL_data_prepared.csv")
```

```r
bal_vars <- c("hh_size","g2002","g2000","p2004","p2002","p2000","sex","age")

d_hh <- d %>% group_by(hh_id,treatment) %>% summarize_at(c(bal_vars,"voted"),mean)

d_hh %>% group_by(treatment) %>% summarize_at(bal_vars,mean) %>%
  gather(key="Variable",value="value",-treatment) %>%
  spread(key="treatment",value="value") %>%
  select(Variable,Control,`Civic Duty`, Hawthorne,Self,Neighbors) %>%
  arrange(factor(Variable,levels=c("hh_size","g2002","g2000","p2004","p2002","p2000","sex","age"))) %>%
  kable(digits=4,booktabs=T)
```

Variable	Control	Civic Duty	Hawthorne	Self	Neighbors
hh_size	1.9124	1.9108	1.9100	1.9109	1.9100
g2002	0.8344	0.8361	0.8357	0.8352	0.8351
g2000	0.8663	0.8654	0.8668	0.8625	0.8653
p2004	0.4166	0.4155	0.4188	0.4209	0.4227
p2002	0.4087	0.4099	0.4117	0.4104	0.4060
p2000	0.2649	0.2664	0.2632	0.2629	0.2631
sex	0.5023	0.5027	0.5032	0.5014	0.5046
age	51.3140	51.1790	51.2041	51.2442	51.3423

Balance is extremely good so randomization was probably very successfully implemented.

5. GGL don't bother to do a t-test for the difference-in-means, but let's do it ourselves. Conduct a t-test for the difference in mean household age between the Control and 'Neighbors' conditions. Interpret the result.

```r
d_hh %>% ungroup() %>%
  filter(treatment %in% c("Control","Neighbors")) %>%
  t.test(age~treatment, data=.)
```

```
##
##  Welch Two Sample t-test
##
## data:  age by treatment
## t = -0.28124, df = 28381, p-value = 0.7785
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.2258195  0.1691470
## sample estimates:
##   mean in group Control mean in group Neighbors
##                51.31400                51.34233
```

We cannot reject the null hypothesis that the two groups have the same mean.

6. Now let's look at the results of the experiment. Perform a simple difference-in-means t-test for voter turnout between the Control and 'Neighbors' groups in the individual data. Interpret the result.

```r
d %>% filter(treatment %in% c("Control","Neighbors")) %>%
  t.test(voted~treatment,data=.)
```

```
##
##  Welch Two Sample t-test
##
## data:  voted by treatment
```

```
## t = -30.207, df = 52613, p-value < 2.2e-16
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.08658577 -0.07603405
## sample estimates:
##   mean in group Control mean in group Neighbors
##               0.2966383               0.3779482
```

The 8% points difference in mean turnout is highly statistically significant so we can reject the null hypothesis that the difference in means is equal.

7. Now run an OLS regression to understand the effect of each treatment on voter turnout, to replicate column (a) of Table 3. (If you prefer you can use treatment as a factor variable, not a series of dummies like the authors use. For this question do not adjust the standard errors). Interpret the results in terms of the change in probability of voting in the election.

```
d <- d %>% mutate(treatment=factor(treatment,levels=c("Control","Civic Duty","Hawthorne","Self","Neighbo

d %>%  zelig(voted~treatment,data=., model="ls", cite=F) %>%
  from_zelig_model() %>%
  stargazer(digits=4, header=F, keep.stat=c("n"))
```

Table 2:

	Dependent variable:
	voted
treatmentCivic Duty	0.0179***
	(0.0026)
treatmentHawthorne	0.0257***
	(0.0026)
treatmentSelf	0.0485***
	(0.0026)
treatmentNeighbors	0.0813***
	(0.0026)
Constant	0.2966***
	(0.0011)
Observations	344,084
Note:	*p<0.1; **p<0.05; ***p<0.01

In the control condition, average turnout is 29.7%. The Civic Duty treatment increases this by 1.8%, the Hawthorne treatment by 2.6%, the Self treatment by 4.9% and the Neighbors treatment by 8.1%..

9. Repeat your regression but this time with standard errors clustered to the household level. What difference does this make? Why do we do this?

```
#d %>% lm_robust(voted~treatment, data=., clusters=hh_id) %>%
#  tidy() %>%
#  select(term,estimate,std.error, p.value) %>%
#  kable(digits=4)%>%
```

```r
#  kable_styling(bootstrap_options = c("striped", "hover"))

# An alternative function for low-memory computers
library(miceadds)

d %>% lm.cluster(voted~treatment, cluster="hh_id") %>%
  summary() %>%
  xtable(digits=5)
```

R^2= 0.00339

```
Estimate   Std. Error     t value       Pr(>|t|)
```

(Intercept) 0.29663831 0.001309562 226.517174 0.000000e+00 treatmentCivic Duty 0.01789934 0.003236627 5.530247 3.197805e-08 treatmentHawthorne 0.02573631 0.003257893 7.899680 2.796207e-15 treatmentSelf 0.04851319 0.003300020 14.700879 6.362799e-49 treatmentNeighbors 0.08130991 0.003369555 24.130757 1.189084e-128 % latex table generated in R 3.5.2 by xtable 1.8-3 package % Thu Apr 11 21:45:56 2019

	Estimate	Std. Error	t value	Pr(>|t|)
(Intercept)	0.29664	0.00131	226.51717	0.00000
treatmentCivic Duty	0.01790	0.00324	5.53025	0.00000
treatmentHawthorne	0.02574	0.00326	7.89968	0.00000
treatmentSelf	0.04851	0.00330	14.70088	0.00000
treatmentNeighbors	0.08131	0.00337	24.13076	0.00000

The standard errors are now larger, reflecting the fact that treatment was applied at the household-level, so the observed outcomes for individuals in the same household are likely to be highly correlated, reducing the effective number of independent observations in our data and increasing our uncertainty about the true size of the effects.

10. Next, we want to add block-level fixed effects to our model to reproduce column (b) of Table 3. However, there are many (10,000) blocks so your computer probably doesn't have enough memory to run this regression directly. An equivalent methodology is to remove the between-group variation in the treatment variable manually and then run the same regression as in Q8. To do this for the 'Neighbors' treatment:
1. Create a dummy variable for individuals that received the 'Neighbors' treatment,
2. Remove the other treatments from your dataset (so you are left with just 'Neighbors' and 'Control' units),
3. For each 'cluster' group calculate the mean value of the binary 'Neighbors' treatment variable,
4. Subtract the cluster mean from the individual values of the Neighbors treatment variable.
5. Run the same regression as in Q8 but using the cluster-mean-centered treatment variable you just created as the explanatory variable.

How do the results change? How does this change the comparisons we are making in the regression?

```r
d_block <- d %>% mutate(Neighbors=ifelse(treatment=="Neighbors",1,0)) %>%
  group_by(cluster) %>%
  mutate(Neighbors_cluster_avg=mean(Neighbors,na.rm=T)) %>%
  ungroup() %>%
  mutate(Neighbors=Neighbors-Neighbors_cluster_avg)

d_block %>% filter(treatment %in% c("Neighbors","Control")) %>%
  zelig(voted~Neighbors,data=.,model="ls", cite=F) %>%  from_zelig_model() %>%
  stargazer(digits=4, header=F, keep.stat=c("n"))
```

Our comparisons are now *within* each block. So we are comparing only treated and control individuals within each block.

Table 3:

	Dependent variable:
	voted
Neighbors	0.0822***
	(0.0026)
Constant	0.3057***
	(0.0010)
Observations	229,444
Note:	*p<0.1; **p<0.05; ***p<0.01

11. Add covariates (g2000,g2002,p2000,p2002,p2004) to your model from Q10 to reproduce column 3 of Table 3. How do the results change when we add covariates?

```
d_block %>% filter(treatment %in% c("Neighbors","Control")) %>%
  zelig(voted~Neighbors + g2000 + g2002 + p2000 + p2002 + p2004,data=., model="ls", cite=F) %>%
  from_zelig_model() %>%
  stargazer(digits=4, header=F, keep.stat=c("n"))
```

Table 4:

	Dependent variable:
	voted
Neighbors	0.0815***
	(0.0025)
g2000	−0.0048*
	(0.0028)
g2002	0.0988***
	(0.0027)
p2000	0.0990***
	(0.0022)
p2002	0.1342***
	(0.0020)
p2004	0.1551***
	(0.0019)
Constant	0.0902***
	(0.0027)
Observations	229,444
Note:	*p<0.1; **p<0.05; ***p<0.01

Adding covariates explains more of the variation in turnout which is not explained by treatment. This makes only a slight difference to our coefficient estimates (adjusting just for the very tiny imbalance that remains between treatment groups after randomization), but systematically reduces the standard errors of our treatment coefficients. This occurs because with more of the variation in the outcome accounted for, the remaining variation in the outcome can be more confidently attributed to treatment (and not something else).

12. In place of an OLS regression, use a logit regression model to run the same model as in Q8. How would you interpret the results?

```
d %>% zelig(voted~treatment,data=., model="logit", cite=F) %>%
  from_zelig_model() %>%
  stargazer(digits=4, header=F, keep.stat=c("n"))
```

Table 5:

	Dependent variable:
	voted
treatmentCivic Duty	0.0844^{***}
	(0.0121)
treatmentHawthorne	0.1205^{***}
	(0.0120)
treatmentSelf	0.2229^{***}
	(0.0119)
treatmentNeighbors	0.3651^{***}
	(0.0117)
Constant	-0.8634^{***}
	(0.0050)
Observations	344,084
Note:	*p<0.1; **p<0.05; ***p<0.01

```
coefs_logit <- d %>% zelig(voted~treatment,data=., model="logit", cite=F) %>%
  from_zelig_model() %>%
  tidy() %>%
  select(term,estimate)
```

We can either interpret the coefficients as a β change in the log-odds of voting, or - more intuitively - as a $100 * (exp^{\beta} - 1)\%$ change in the odds (relative probability) of voting. So, for example, for the Neighbors treatment, there is an estimated 44.06% increase in the odds of voting compared to the control condition.

This makes sense: from our OLS regression the control turnout is 29.7% so the *odds* of turnout under control are $\frac{0.297}{(1-0.297)}$ =0.4224751. Under the Neighbors treatment the turnout is 37.8%, so the odds of turnout are $\frac{0.378}{(1-0.378)}$ =0.607717. The *increase in the odds* from control to treatment is therefore $\frac{0.608-0.422}{0.422} = 44.08\%$, the same as suggested by the odds ratio from our logit regression.

If we wanted to return to the original probabilities directly from our logit regression estimates we need to go a step further. We need to take the inverse-logit ($\frac{exp(\beta)}{(1+exp(\beta))}$) of the expected values of the regression in the treated condition ($\frac{exp(-0.85+0.36)}{(1+exp(-0.85+0.36))}$) and subtract the inverse-logit of the expected values of the regression in the control condition ($\frac{exp(-0.85)}{(1+exp(-0.85))}$). So moving from control to treatment increases voting by

$\frac{exp(-0.85+0.36)}{(1+exp(-0.85+0.36))} - \frac{exp(-0.85)}{(1+exp(-0.85))} = 0.0805$ % points.

13. Predict the first difference (the change) of the probability of voting when moving from the 'Control' to 'Neighbors' treatment category using your logit model from Q12. ***Hint: Use Zelig in R and Clarify in Stata***

```
fd_mean <- d %>% mutate(Neighbors=ifelse(treatment=="Neighbors",1,0)) %>%
  filter(treatment %in% c("Neighbors","Control")) %>%
  zelig(voted~Neighbors,data=., model="logit", cite=F) %>%
  setx(Neighbors=0) %>%
  setx1(Neighbors=1) %>%
  sim(num=10000) %>%
  zelig_qi_to_df() %>%
  group_by(setx_value) %>%
  summarize(fds=mean(expected_value)) %>%
  spread(key="setx_value",value="fds") %>%
  mutate(diff=x1-x) %>%
  pull(diff)
```

The first differences estimation suggests a mean first difference in turnout between control and treatment of 8.133% points.

14. How does the data processing the authors conduct on p.36-37 (under 'Study Population') affect your interpretation of the conclusions?

They remove any household with a hard-to-find address, or where there are many apartment buildings, so they are probably under-sampling from poorer neighbourhoods. Perhaps the effects would be different in this population. They also exclude absentee ballot voters and people who did not vote in 2004 so they are focusing on only the people most likely to be affected by their treatment. That means the percentage gains in turnout we see would probably not be true across the whole electorate.

15. How generalizable are the findings of this study to other elections? To the same set of elections in 2010? To neighbouring Indiana in the same year? To elections in Brazil?

Strictly, we cannot generalize beyond the population from which our random sample was drawn: the August 2006 primary election in Michigan. However, the degree to which treatment effects will differ depends on how different the context is. So if the political and economic circumstances were similar in 2010 our estimates may be reasonable. They may be a bit worse in a different state with potentially a different political culture. And they probably make very little sense in a country like Brazil that has compulsory voting.