Making a Difference: The Consequences of Electoral Experiments
Codebook

Tara Slough*

December 10, 2023

This codebook describes the variables in the original raw data and the processed datasets included in this replication package. Note that descriptions of the variables in the raw datasets used (electoral data) and replication datasets is available following the links in the README. All datasets described here are located in the `results` folder in the CodeOcean capsule unless otherwise noted.

studies_maei.csv

`studies_maei.csv` (located in `data/raw/`) contains one dataset necessary to construct Figure 1. The data consists of original back-of-the-envelope calculations described in Table A8. See Table C4 for a description of the variables.

colorado.Rdata

`colorado.Rdata` contains the four dataframes necessary to conduct the simulation reported in Figures 2-3 and A11. The dataframe `dat` contains panel data on electoral results in Colorado State House elections from 2010-2018 (see Table C1). The dataframe `sh2016p` contains precinct-level Colorado State House results from 2016 and registration data from 2018 (see Table C2). The dataframe `ush2016p` contains precinct-level US House results from 2016 and registration data from 2018 (see Table C3). The dataset `indiv` comes directly from Morris (2018).

pap_data.Rdata

`pap_data.Rdata` contains the dataframe `trials` with data from the EGAP and AEA registries. See Table C5 for description of variables.

gg.Rdata

`gg.Rdata` contains the dataframe `gg_electoral`, which contains the variables used in the Gerber and Green (2000) application. See Table C6 for description of variables.

bhm.Rdata

`bhm.Rdata` contains the dataframes `bhm`, `bhm_n`, and `panel` which contains replication data from Boas, Hidalgo, and Melo (2019) and relevant electoral data from Brazil's Tribunal Superior Eleitoral (TSE). The dataframe `bhm` comes directly from the Boas, Hidalgo, and Melo (2019) replication file, for which there is an existing codebook. See Table C7 for description of the variables in `bhm_n` and Table C8 for description of the variables in `panel`.

*Assistant Professor, New York University

Variable	Description	Source	N	Minimum	Maximum
year	Election year	CO Secretary of State	325	2010	2018
District	State House District Number	CO Secretary of State	325	1	65
DEM	Votes for Democratic Party candidate	CO Secretary of State	325	0	40,121
REP	Votes for Republican Party candidate	CO Secretary of State	325	0	40,011
total	Total two-party votes	CO Secretary of State	325	0	57,179
vote_marg	Votes for Democratic candidate - votes for Republican candidate	Calculated from CO Secretary of State	325	-36,060	40,121
dem_marg	Proportion of registered Democrats - proportion of registered Republicans (active or inactive)	Calculated from CO Secretary of State	325	-0.607	0.710
Atotal	Number of active registered voters	CO Secretary of State	325	11,844	73,236
TOTAL	Number of registered voters	CO Secretary of State	325	16,702	81,775
vote_prop	Democratic margin of victory (or defeat) as a share of registered voters (vote_marg/TOTAL)	CO Secretary of State	325	-0.733	0.592
lag	Lagged vote_prop	CO Secretary of State	260	-o.733	0.592

Table C1: Variables in the dat dataframe. N gives the number of non-missing observations.

Variable	Description	Source	N	Minimum	Maximum
Precinct	Unique precinct identifier	CO Secretary of State	2,990	—	—
Total	Total two-party votes	CO Secretary of State	2,990	0	2,354
REP	Votes for Republican Party candidate	CO Secretary of State	2,990	0	1,568
DEM	Votes for Democratic Party candidate	CO Secretary of State	2,990	0	1,748
District	State House District Number	CO Secretary of State	2,990	1	65
year	Election year (2016)	CO Secretary of State	2,990	2016	2016
COUNTY	County in which precinct is located	CO Secretary of State	2,990	—	—
ACTIVE.	Number of active registered voters (2018)	CO Secretary of State	1	4,693	73,236
INACTIVE	Number of inactive registered voters (2018)	CO Secretary of State	2,990	1	4,003
TOTAL	Number of registered voters (2018)	CO Secretary of State	2,990	1	5,075
ratio	Democratic votes (2016) as a proportion of registered voters (2018)	Calculated from CO Secretary of State	2,990	0	1.384
b1	ratio×TOTAL (equivalent to DEM)	Calculated from CO Secretary of State	2,990	0	1,748
b2	(1-ratio)×TOTAL	Calculated from CO Secretary of State	2,990	-323	4,504

Table C2: Variables in the sh2016 dataframe. N gives the number of non-missing observations.

Variable	Description	Source	N	Minimum	Maximum
`Precinct`	Unique precinct identifier	CO Secretary of State	2,990	—	—
`Total`	Total two-party votes	CO Secretary of State	2,990	0	2,448
`REP`	Votes for Republican Party candidate	CO Secretary of State	2,990	0	1,566
`DEM`	Votes for Democratic Party candidate	CO Secretary of State	2,990	0	1,614
`District`	State House District Number	CO Secretary of State	2,990	1	7
`year`	Election year (2016)	CO Secretary of State	2,990	2016	2016
`COUNTY`	County in which precinct is located	CO Secretary of State	2,990	—	—
`ACTIVE.`	Number of active registered voters (2018)	CO Secretary of State	1	4,693	73,236
`INACTIVE`	Number of inactive registered voters (2018)	CO Secretary of State	2,990	1	4,003
`TOTAL`	Number of registered voters (2018)	CO Secretary of State	2,990	1	5,075
`ratio`	Democratic votes (2016) as a proportion of registered voters (2018)	Calculated from CO Secretary of State	2,990	0	1.374
b1	`ratio`×`TOTAL` (equivalent to `DEM`)	Calculated from CO Secretary of State	2,990	0	1,614
b2	(1-`ratio`)×`TOTAL`	Calculated from CO Secretary of State	2,990	-315	4,508

Table C3: Variables in the `ush2016` dataframe. N gives the number of non-missing observations.

Variable	Description	Source	N	Minimum	Maximum
`Study`	Index for number of studies (ranked by lower bound of average $MAEI_d$)	Original	1	6	
`lwr`	Lower bound of average $MAEI_d$ when $E[a_c(0)] = \frac{1}{2} \forall c$ if cluster-randomized. ($E[a_c(0)] = 1$ if individually randomized.)	Original	0.008	0.279	
`lwr`	Upper bound of average $MAEI_d$ when $E[a_c(0)] = 1 \forall c$ if cluster-randomized. ($E[a_c(0)] = 1$ if individually randomized.)	Original	0.008	0.594	
`type`	Indicates whether interval bounds the maximum or average $MAEI_d$.	Original	—	—	
`rand_type`	Indicates whether treatment is assigned to individuals or clusters consisting of > 1 voter.	Original	—	—	

Table C4: Variables in the `studies_maei.csv` dataset. N gives the number of non-missing observations.

Variable	Description	Source	N	Minimum	Maximum
`Treatment`	Classification of intervention	Original coding of EGAP and AEA registries	129	—	—
`Country`	Country where intervention was conducted	EGAP and AEA registries	129	—	—
`Election`	Type of election in which intervention was conducted	Original coding of EGAP and AEA registries	129	—	—
`Office`	Office targeted by intervention	Original coding of EGAP and AEA registries	129	—	—
`Start`	Start of intervention	EGAP and AEA registries	129	—	—
`IRB`	Page number of discussion of IRB (if pre-analysis plan accessible) or indication of IRB approval on the registry	Original coding of EGAP and AEA registries	63	—	—
`Ethics_not_IRB`	Page number of discussion of ethical considerations beyond IRB approval in pre-analysis plan	Original coding of EGAP and AEA registries	107	—	—
`Outcomes`	Indicator for discussion of aggregate electoral impact/election outcomes as an ethical concern among studies with ethics discussion	Original coding of EGAP and AEA registries	6	0	1
`dataset`	Indicator for registry in which study was located. (Defaults to EGAP for cross-listed studies)	EGAP and AEA registries	129	—	—
`Registered`	Date registered in registry	EGAP and AEA registries	129	2012-10-12	2020-11-26
`region`	Region (continent) in which study was conducted	EGAP and AEA registries	129	—	—
`trt`	Classification of intervention	Original coding of EGAP and AEA registries	129	—	—
`start_date`	Date of start of intervention	EGAP and AEA registries	129	1998-10-31	2020-10-29
`yr`	Year of start of intervention	EGAP and AEA registries	129	1998	2020
`US`	Indicator for experiments conducted in US	EGAP and AEA registries	129	0	1

Table C5: Variables in the `trials` dataframe. N gives the number of non-missing observations.

Variable	Description	Source	N	Minimum	Maximum
`Race`	Office on 1998 ballot in New Haven	Connecticut Secretary of State electoral results	18	chr	chr
`uncontested`	Denotes whether race was contested or uncontested	Connecticut Secretary of State electoral results	18	chr	chr
`pred`	Predicted margin to pivotality	Calculated from 10/9-12 Mason-Dixon poll as reported in Cook Political Report (1998)	2	0.089	0.103
`psid`	Ex-post margin to pivotality (as a share of registered voters)	Connecticut Secretary of State electoral results	18	0.001	0.303
`maeid`	Maximum Aggregate Electoral Impact ($MAEI_d$) calculated for each race. Note that this quantity is missing in districts that contain only a portion of New Haven (the state legislative races).	Calculated from Gerber and Green (2000) replication data and Connecticut Secretary of State electoral results.	9	0.010	0.371

Table C6: Variables in the `gg_electoral` dataframe. N gives the number of non-missing observations.

Variable	Description	Source	N	Minimum	Maximum
ibge7_code	Municipal identifier (7 digits) from Brazilian Institute of Geography and Statistics (IBGE)	IBGE	46	–	–
n	Number of subjects in experiment	Boas, Hidalgo, and Melo (2019) replication data	46	416	chr
nt	Number of treated subjects	Boas, Hidalgo, and Melo (2019) replication data	46	25	279
ANO_ELEICAO	Election year	TSE	46	2016	2016
SG_UF	State	TSE	46	–	–
CD_MUNICIPIO	TSE municipal code	TSE	46	–	–
tot	Total registered voters in 2016	TSE	46	7,038	121,877
tse_code	TSE municipal code	TSE	46	–	–
ibge7_code	IBGE 6-digit municipal identifier	IBGE	46	–	–
mesoregion	IBGE mesoregion identifier	IBGE	46	–	–
microregion	IBGE microregion identifier	IBGE	46	–	–
UF	State	IBGE	46	–	–

Table C7: Variables in the bhm_n dataframe. N gives the number of non-missing observations.

Variable	Description	Source	N	Minimum	Maximum
`tse_code`	TSE municipal code	TSE	736	–	–
`dif`	Difference in votes between top-2 candidates in election t	TSE	736	1	22,607
`nvotes`	Total votes in election t	TSE	736	2,552	1,735,336
`ncand`	Number of mayoral candidates in election t	TSE	736	1	140
`year`	Election year t	TSE	736	2004	2016
`ANO_ELEICAO`	Election year of experiment	TSE	736	2016	2016
`SG_UF`	State	TSE	736	–	–
`CD_MUNICIPIO`	TSE municipal code	TSE	736	–	–
`tot`	Total registered voters in 2016	TSE	46	3,714	1,119,271
`lead_dif`	Difference in votes between top-2 candidates in election $t+1$	TSE	552	1	22,607
`pred`	Predicted difference in vote share between top-2 candidates as a share of registered voters	Constructed from TSE data	552	0.088	0.222
`psid`	Fifth percentile of predictive interval, $\underline{\psi}_d$	Constructed from TSE data	552	-0.023	0.111

Table C8: Variables in the `panel` dataframe. N gives the number of non-missing observations. The sample is all municipalities in Pernambuco.

Codebook: References

Boas, Taylor, F. Daniel Hidalgo, and Marcus André Melo. 2019. "Norms versus Action: Why Voters Fail to Sanction Malfeasance in Brazil." *American Journal of Political Science* 63 (2): 385–400.

Gerber, Alan S., and Donald P. Green. 2000. "The Effects of Canvassing, Telephone Calls, and Direct Mail on Voter Turnout: A Field Experiment." *American Political Science Review* 94 (3): 653–663.

Morris, G. Elliott. 2018. "2018 U.S. House Midterm Elections Forecast." Available at `https://www.thecrosstab.com/project/2018-midterms-forecast/`.