INSTRUCTIONS FOR USE

Lucerne Elections

The document aims to facilitate the use of the online pre- and post-election survey datasets from Lucerne. Please note that the Lucerne elections were held on April 3, 2011. The preelection survey was conducted between March 24-April 2, and the post-election survey between April 4-April 9.

There are two main datasets that researchers can use:

- Original 'uncleaned' datasets.
  - 
    MEDW Switz Wave1-Original _Uncleaned_Lucerne_July_2011.dta
- Cleaned data, which aims to facilitate the use of data.
  - 
    MEDW Switz Wave1-Lucerne_cleaned_July6_2011.dta

This document describes some unique features of the 'uncleaned' dataset and records what adjustments have been made between the 'clean data' and the original data. We have also posted 'do-file' that generates 'clean data', which may help all users to be familiar with the dataset. If you have any suggestions regarding the 'clean data', please let us know.

UNCLEANED DATA

1. Time Stamps

We would like to remind users of the availability of *time stamp* variables in case they would like to use them in their analysis. Please note that in the cleaned data, we erased all time-stamp.

HOW TO USE "TIME STAMPS" VARIABLES
(From Harris/Decima Technical Reports)

Time Stamps: The way that timing variables work in this study is that they each capture how long the survey has taken (in seconds) at the point the respondent crosses the variable. To analyze how long someone has taken between two timestamps, subtract the first time stamp variable from the second and this will represent how long it took the respondent to get from the first variable to the second. In order to convert to minutes, divide the outcome by 60.

HOW TO CALCULATE THE DURATION OF SURVEY

One can use two variables to calculate the approximate length of the survey for respondents. These variables are "tim090" "tim001" for the pre-election survey and "tim128 and tim091" for the post-election survey.

2. Experiments

a. The pre- and post-election surveys have *experiments* with regard to the 'satisfaction with democracy' question at four administrative levels: Municipality, canton (Zurich), federal level and the EU. Half of the respondents got satisfaction with democracy question at the beginning of the survey and half of them got at the end of the survey in the pre-election and post-election survey. These are labelled as questions **q1a1 q1a2 q1a3 q1a4** AND **q1a1_1 q1a2_1 q1a3_1 q1a4_1** for preelection survey and **pq1a1 pq1a2 pq1a3 pq1a4** AND **pq1a1_1 pq1a2_1 pq1a3_1 pq1a4_1.**

b. *Only for post-election survey data*: For the turnout question, there was an experiment with the response options provided. Half of the sample was asked **pq5_1** and the other half **pq5_2.**

3. Weighting

Harris/Decima created different weights for the data which are described in the technical report. The relevant sections are copied below.

(From Harris/Decima Technical Reports)

At the conclusion of the data collection and cleaning, Harris/Decima weighted the data by each quota stratum to reflect the actual proportions found in the population. This ensures the findings from the research can be extrapolated to the entire population with accuracy. Harris/Decima uses a standard procedure for calculating weighting factors, based on established methodological standards and extensive experience in sample weighting over literally hundreds of projects. This procedure involves calculating the actual population within each segment and the true proportion of the sample that would fall into each segment if the survey were conducted on strictly a random basis. Into this number is divided the actual segment sub-sample to produce a weighting factor that is then used to "weight" the data for that segment. While there are various ways of accomplishing this task, this procedure is the most straightforward and effective.

The datafile includes eight weights, each of which was created to suit a unique purpose and some of which were used to create other weights:

Name	Factors	Use for Pre/Post	
PRE_WEIGHT1	age, gender, education	Pre	
PRE_WEIGHT2	age, gender, education and likelihood to vote (vote turnout)	Pre	

PRE_WEIGHT3	age, gender, education, likelihood to vote (vote turnout) and vote intention (actual election results)	Pre	
PRE_WEIGHT4	age, gender, education and vote intention (actual election results)	Pre	
POST_WEIGHT1	age, gender, education	Post	
POST_WEIGHT2	age, gender, education and likelihood to vote (vote turnout)	Post	
POST_WEIGHT3	age, gender, education, likelihood to vote (vote turnout) and vote intention (actual election results)	Post	
POST_WEIGHT4	age, gender, education and vote intention (actual election results)	Post	

CLEANED DATA

1. Weighting

The same weights are available in the clean dataset. No changes were made.

2. How To Distinguish Pre and PostElection Surveys:

Two new variables are created so that users can easily choose whether they would like to work on the pre-election survey data or the post-election survey data only. These variables and their description are as follows:

- *precompletes*: those who completed the pre-election survey
- *postcompletes*: Those who completed the post-election survey

Note: 'Section' variable has more detailed information about the pre and postelection surveys.

Most Frequently Used Variables:

Some frequently used variables such as age, education, income, turnout and others have been generated so that users can easily identify them.

Variables:

Following variables were created so that researchers can easily utilize most-frequently variables. Please note that 'do not know answers' were coded as missing (=.). Please use 'do-file' to see how these variables are created.

- **turnout:** Using two experiments for turnout, a single turnout variable was created **(pq5_1==4 and pq5_2==1**). For these questions, those who answers do not know or do not respond are coded as missing (=.).

- **satisfaction with democracy (swd):** Using two experiments for satisfaction with democracy, new variables were created. Those who answers do not know are coded as missing (=.).

  For Preelection
  Swdmunicpre: SWD in Municipality
  Swdcantonpre: SWD in Canton of Lucerne
  Swdfederalpre: SWD in Switzerland
  swdEUpre: SWD in EU

  For Postelection
  Swdmunicpost: SWD in Municipality
  Swdcantonpost: SWD in Canton of Lucerne
  swdfederalpost : SWD in Switzerland
  swdEUpost: SWD in EU

- **female**: q268 was recoded and 'female' variable was created. 1=woman and 0=man.
- **age**: continuous variable (18-79), and this variable was not changed.
- **education**: education variable was generated using original variable of sd2a. No recoding was done so that users can decide how to use this education variable.
- **income**: income variable was generated, and no recoding was made. Original variable, sd5 was also kept in the dataset.
- **townsize**: townsize variable was generated, and no recoding was made. Original variable, sd7, was not changed.
- **vote**: This variable was created following such coding:
  If one marks the party list (a response to *pq6a1*), AND gives the name of the party (*pq6a3*), the name of the party was coded.
  If one marked empty list (*pq6a1*), they were asked whether they put the party name on the list or not (*pq6a2*). If the party name was put on (*pq6a3*), the party name was coded.
  If one marked empty list (pq6a1) and marked 'No' to putting a party name on the list (*pq6a2*), then respondents did a 'distribution of votes' for parties (**pq6b2a1- pq6b2a30**). The name of the party that got most votes was chosen. In case of ties, we chose randomly the name of the party.
  Here is the coding of the parties for vote variable:

SVP=3	**Grune**=5	**Gruneliberal**=7
SP=4	**CVP**=1	**Others**=88
FDP=2	**EVP**=6	**99=Do Not Know**

3. Adjustments

The Clean dataset excludes those who fall into the 'neither section started' and 'prelection incomplete' categories.

Note: Section=8 ("neither section started") and Section=2 ("pre-election incomplete") were dropped.

4. Failures

Flag : Indicates a potential non-logic in the survey. Flags those who answered Q6b1_1 = 1 (Ja) and Q6b1_3 = 2 (Nein) for Lucerne pre-election study as "illogic at Q6b1_1 and Q6b1_3."

5. In-Survey Quality (ISQ) Measures

The dataset has a number of variables to indicate ISQ measures. These are labeled pre_grid, pre_resp, pre_time, post_grid, post_resp, and post_time.

There are also summary measures: **pre_fail** and **post_fail** indicate whether any were failed (yes/no), while **PRE_FAIL2** and **POST_FAIL2** indicate how many were failed.

The "**grid**" failure has to do with straightlining. There is some concern that this could have been done legitimately. We did **NOT** exclude them and leave the decision to researchers.

The "**resp**" failure indicates those who did not click "4" when directed to do so. pre_resp for preelection survey and post_resp (postelection survey) were **DELETED**.

The "**time**" failure indicates those who did not take long enough to do the study. These observations with pre_time and post_time were **DELETED** from the file.