Replication materials for 'Worth Weighting? How to Think About and Use Weights in Survey Experiments'

November 16, 2017

This folder contains replication materials for "Worth Weighting? How to Think About and Use Weights in Survey Experiments" by Miratrix, Luke W. and Sekhon, Jasjeet S. and Theodoridis, Alexander G. and Campos, Luis F.

Contents

Simulation Code:

- `run_paper_simulations.R`: This file contains code to run the simulations presented in Section 5 of "Worth Weighting? How to Think About and Use Sample Weights in Survey Experiments." See the *Usage* section below for instructions on running this as well as approximate runtimes.

Data Analysis Code:

- `run_paper_data_analysis.R`: This file contains code to generate the analyses that are presented in Section 6 of "Worth Weighting? How to Think About and Use Sample Weights in Survey Experiments". It will generate a series of output files for the analyzed experiments. See the *Usage* section below for instructions on running this as well as approximate runtimes.
- `run_paper_analysis_noPS.R`: This file is a version of the prior file, except it does not post-stratify on party ID. See the *Usage* section below for instructions on running this as well as approximate runtimes.
- `run_paper_data_analysis_figures.R`: This file contains code to visualize the output from the prior files, producing the results presented in Section 6 of "Worth Weighting?" This code requires all of the saved output from both "`run_paper_data_analysis.R`" and "`run_paper_analysis_noPS.R`". See the *Usage* section below for instructions on running this as well as approximate runtimes.

Data:

- `Survey_Experiment_Data_table.csv`: Tabular format of Survey_Experiment_Data.Rda for viewing purposes only.
- `Survey_Experiment_Data.Rda`: This file contains the raw survey experiment data. It is used by "`run_paper_data_analysis.R`"

Support Code:

`-_function_analysis.R`: contains wrapper code for running one analysis on a given experiment. `-_paper_analysis_labels.R`: contains labels for experiment names and subsets for analysis (this is basically a datafile). `-_simulation_functions.R`: contains functions to run and analyze the simulation. `-_survey_exp_toolkit.R`: contains generic functions to implement the various estimators for survey experiments considered in the paper.

Reference Citation:

- `DOI-10.7910-DVN_52UGJT.bib`: contains .bib reference for these replication materials.

Usage:

The above files can be run in R directly or, more easily, they could be run in a Terminal. After changing directories to the directory containing all the files, you can run the analysis, figures and simulation files with the following commands in Terminal. Below the commands are the run times (in seconds). Note that the simulations take approximately 8 hrs. The analysis files should take under 1 hour, total. For exact replicability of the results presented in the paper, please see the next section.

Run the Analysis of Survey Experiments:

```
R CMD BATCH --no-save --no-restore ./run_paper_data_analysis.R ./run_paper_data_analysis.out
```

```
Runtime:
    user   system  elapsed
1925.550   24.326 2134.558
```

Run the Analysis of Survey Experiments (with no post-stratification on Party ID):

```
R CMD BATCH --no-save --no-restore ./run_paper_data_analysis_noPS.R
  ./run_paper_data_analysis_noPS.out
```

```
Runtime:
    user   system  elapsed
1194.653   18.671 1384.403
```

Summarize Analysis of Survey Experiments (make figures, calculate results used in paper):

```
R CMD BATCH --no-save --no-restore ./run_paper_data_analysis_figures.R
  ./run_paper_data_analysis_figures.out
```

```
Runtime:
 user  system elapsed
0.363   0.050   0.431
```

Run the Simulation Studies:

```
R CMD BATCH --no-save --no-restore ./run_paper_simulations.R ./run_paper_simulations.out
```

```
Runtime:
     user    system   elapsed
29634.915     1.793 29658.523
```

On Windows, you can use similar commands by following instructions found here.

Replicability/Session Info:

We do expect differences when running this code on different systems (and with different seeds), but the differences all tend to be quite small and the general trends remain. If you would like to replicate the exact numbers and figures used in the article, below you will find the session info. Please install and load all package versions (including the R version) to ensure the exact replication.

```
sessionInfo()
```

```
R version 3.3.0 (2016-05-03)
Platform: x86_64-apple-darwin13.4.0 (64-bit)
Running under: OS X 10.13.1 (unknown)

locale:
[1] en_US.UTF-8/en_US.UTF-8/en_US.UTF-8/C/en_US.UTF-8/en_US.UTF-8

attached base packages:
[1] stats     graphics  grDevices utils     datasets  methods   base

other attached packages:
[1] extrafont_0.17 astsa_1.7      car_2.1-4      Matching_4.9-2 MASS_7.3-45
[6] reshape2_1.4.2 xtable_1.8-2   foreign_0.8-67 plyr_1.8.4

loaded via a namespace (and not attached):
 [1] Rcpp_0.12.12       Rttf2pt1_1.3.4     magrittr_1.5       splines_3.3.0
 [5] lattice_0.20-34    minqa_1.2.4        stringr_1.2.0      tools_3.3.0
 [9] nnet_7.3-12        pbkrtest_0.4-7     parallel_3.3.0     grid_3.3.0
[13] nlme_3.1-131       mgcv_1.8-17        quantreg_5.29      extrafontdb_1.0
[17] MatrixModels_0.4-1 lme4_1.1-12        Matrix_1.2-8       nloptr_1.0.4
[21] stringi_1.1.2      SparseM_1.76
```