

JASA ACS Reproducibility Initiative - Author Contributions Checklist Form

The purpose of the Author Contributions Checklist (ACC) Form is to document the code and data supporting a manuscript, and describe how to reproduce its main results.

As of Sept. 1, 2016, the ACC Form must be included with all new submissions to JASA ACS.

This document is the initial version of the template that will be provided to authors. The JASA Associate Editors for Reproducibility will update this document with more detailed instructions and information about best practices for many of the listed requirements over time.

Data

Abstract (Mandatory)

The data set analyzed in the paper was provided by Ninth Decimals, 625 Ellis St., Ste. 301, Mountain View, CA 94043, a marketing platform using location data, see <http://www.ninthdecimal.com/>.

Availability (Mandatory)

The data will be displayed on DataVerse. Here is the link:

<https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/H4AOQP>.

Description (Mandatory if data available)

The data was made available to the authors in exchange for statistical services. The data set contains activations of applications on mobile phones recorded at the dealerships of a major auto brand in a US state over a 15 months period. The auto brand and the state are not divulged for confidentiality purposes. The data is provided in a capture-recapture format; the first column contains the unique identifier of the application while the other columns contain its capture history, as explained in the manuscript. Two versions of the data set are provided, one is about daily activations and the other gives weekly aggregated data.

Optional Information (complete as necessary)

NA

Code

Abstract (Mandatory)

The R codes provided with the manuscript reproduce all the estimation results presented in Section 6 of the paper.

Description (Mandatory)

The R code (R script files with extension .R) and two test files containing the daily and the weekly data sets are provided in the zip file named codes.

Optional Information (complete as necessary)

The following R packages are necessary to successfully use the codes:

- nleqslv (version 3.3)
- ggplot2 (version 2.2.1)
- Rcapture (version 1.4-2)
- multinomRob (version 1.8-6.1)
- lubridate (version 1.6.0)
- readxl (version 1.0.0)

Instructions for Use

Reproducibility (Mandatory)

In this section, the 12 steps to estimate the capture-recapture parameters discussed in Section 6 and to create figures 2, 3, 4 in the paper are given. The user only needs to execute lines of R codes that are contained in the file 'Estimation_Global_App_Data.R'. The steps are as follow.

1. Unzip the archive codes.zip
2. Open the R file 'Estimation_Global_App_Data.R',
3. Set the work directory as the one containing the unzipped files (.txt and .R files) by modifying line 2 of Estimation_Global_App_Data.R,
4. Install and load the following R-packages:
 - Rcapture
 - ggplot2
 - lubridate
 - multinomRob
 - nleqslv
 - readxl,
5. Source the R code (10 files) for loading the data, for calculating the sufficient statistics and for estimating the capture-recapture parameters
6. Source `figure2.R` to create Figure 2 in the manuscript,
7. Execute the R program Esti_param_Mh_Cp to fit closed population model Mh to the data;
8. Execute the R program Esti_param_jollySeber to fit the Jolly-Seber model to the data;
9. Execute the R program Esti_param_Mh_Rd to fit the robust design model to the data;
10. Source `figure3.R` to create Figure 3 in the manuscript,
11. Execute CR_Bootstrap to get bootstrap variance estimates for the three models fitted in step 7, 8, and 9. The default number of trials in CR_Bootstrap is 100, this can be changed using the variable trials
 - For 1000 trials, the program will approximately take 42 min to run
 - For 500 trials, the program will approximately take 20 min
 - For 100 trials, the program will approximately take 4 min,

For the figures in the Supplementary material, do the following:

12. Source ``figure4.R`` to create Figure 3 in the Supplementary Material
13. Source ``figure5.R`` to create Figure 2 in the Supplementary Material
14. Source ``figure6.R`` to create Figure 1 in the Supplementary Material

For the first part of the simulation study, do the following:

15. Source ``CR_simul_1.R``, ``CR_simul_2.R`` and ``CR_simul_3.R`` to load the needed functions for the Monte Carlo study;
16. Execute the next 17 lines to run the three scenarios (RD, ED1, ED2) and get the corresponding simulation results.

For the second and last part of the simulation, do the following:

17. Source ``CR_simul_3_v2.R`` and ``smartP_simul_LP_1_1_v2`` to load the needed functions for the Monte Carlo study;
18. Execute the next 15 lines to get the simulation results (mean increase, RMSE)
19. Execute the following 15 lines to get the absolute biases for the true relative increase ($\text{inc}=0.5, 0.2, 0.8$).

The code for the numerical example, including bootstrap variance calculations, ran in about 10 minutes on a standard laptop.

The first simulation study ran in 10 hours on a standard desktop while the second one used 16 hours.

Replication (Optional)

The 14 steps can be run and a different data set by uploading the new weekly and daily data on lines 4 and 5 of `app_data.R`

Notes

NA