JASA ACS - Author Contributions Checklist (ACC) Form

This form documents the artifacts associated with the article "A Robust, Differentially Private Randomized Experiment for Evaluating Online Educational Programs With Sensitive Student Data" (i.e., the data and code supporting the computational findings) and describes how to reproduce the findings.

Part 1: Data

☐ This paper does not involve analysis of external data (i.e., no data are used or the only data are generated by the authors via simulation in their code).

☒ I certify that the author(s) of the manuscript have legitimate access to and permission to use the data used in this manuscript.

Abstract

Data collected in the "Evaluation of Online Statistics Courses at the University of Wisconsin-Madison During COVID-19" Robust Private Randomized Control Trial (RP-RCT), as described in Section 4 of the manuscript. This data is (2,0) - differentially private. We have received approval by the Education and Social/Behavioral Science Institutional Review Board (IRB) of the University of Wisconsin - Madison to publicly share this data.

Availability

☒ Data **are** publicly available.
☐ Data **cannot be made** publicly available.

If the data are publicly available, see the *Publicly available data* section. Otherwise, see the *Non-publicly available data* section, below.

Publicly available data

☒ Data are available online at: https://github.com/ManjushaKancharla/A-Robust-Differentially-Private-Randomized-Experiment

☐ Data are available as part of the paper's supplementary material.

☐ Data are publicly available by request, following the process described here:

☐ Data are or will be made available through some other mechanism, described here:

Description

The data ("RPRCT_data.csv") can be found in the GitHub repository listed above.

File format(s)

☒ CSV or other plain text.
☐ Software-specific binary format (.Rda, Python pickle, etc.): pkcle
☐ Standardized binary format (e.g., netCDF, HDF5, etc.):
☐ Other (please specify):

Data dictionary

☐ Provided by authors in the following file(s):
☒ Data file(s) is(are) self-describing (e.g., netCDF files)
☐ Available at the following URL:

Part 2: Code

Abstract

The code can be used to reproduce the results in Table 2 of the manuscript.

Description

Code format(s)

☒ Script files
  ☒ R
  ☐ Python
  ☐ Matlab
  ☐ Other:

Supporting software requirements

Version of primary software used   R version 4.1.1

Libraries and dependencies used by the code   R libraries used by this code: dplyr (Version 1.0.7), forcats (Version 0.5.1), utils (Version 4.1.1) and boot (Version 1.3-28).

Parallelization used

☐ No parallel code used
☒ Multi-core parallelization on a single machine/node
  – Number of cores used: 8
☐ Multi-machine/multi-node parallelization
  – Number of nodes and cores used:

Part 3: Reproducibility workflow

Scope

The provided workflow reproduces:

☐ Any numbers provided in text in the paper
☐ The computational method(s) presented in the paper (i.e., code is provided that implements the method(s))
☐ All tables and figures in the paper
☒ Selected tables and figures in the paper, as explained and justified below:

R code to generate results in Table 2 of manuscript.

Workflow

Location

The workflow is available:

☐ As part of the paper's supplementary material.

☒ In this Git repository: https://github.com/ManjushaKancharla/A-Robust-Differentially-Private-Randomized-Experiment

☐ Other (please specify):

Format(s)

☒ Single master code file

☐ Wrapper (shell) script(s)

☐ Self-contained R Markdown file, Jupyter notebook, or other literate programming approach

☐ Text file (e.g., a readme-style file) that documents workflow

☐ Makefile

☐ Other (more detail in *Instructions* below)

Instructions

Results can be reproduced by running the file "RPRCT_analysis.R".

Expected run-time

Approximate time needed to reproduce the analyses on a standard desktop machine:

☐ < 1 minute

☐ 1-10 minutes

☒ 10-60 minutes

☐ 1-8 hours

☐ > 8 hours

☐ Not feasible to run on a desktop machine, as described here: