Replication Archive for

"Inaccurate Statistical Discrimination: An Identification Problem"

J. Aislinn Bohren, Kareem Haggag, Alex Imas, Devin Pope

Review of Economics and Statistics

Overview

The analysis in this paper relies on data from two sets of experiments run on Amazon Mechanical Turk over February 26, 2018 (Survey 1: Workers) and February 27, 2018 (Survey 2: Employers). We provide the Qualtrics raw data (and QSF files), intermediate datasets and R file, and the final cleaning and analysis Stata do file and dataset.

This archive contains five directories:

1. `Scripts`: This directory contains an R file (`1-Process-Part-1.R`) used to lightly process the results from the first worker Qualtrics survey and produce the profiles used in the second employer Qualtrics survey. The directory also includes the primary Stata file (`inaccurate_discrimination.do`) used for the cleaning and all the analysis reported in the main paper and online appendix.

2. `Data-Raw`: This directory contains the raw Qualtrics survey response output for Survey 1 (`Qualtrics-Part1.csv`) and Survey 2 (`Qualtrics-Part2.csv`) and survey payments made to the participants. Amazon Mechanical Turk worker IDs have been removed from the raw files prior to posting.

3. `Data-Extract`: This directory contains both the final analysis file (`inaccurate_discrimination.dta`) and intermediate files produced to

4. `Qualtrics`: This directory includes the QSF files from both survey experiments, as well as PDFs showing the survey flows and all questions.

5. `Results`: This directory includes folders contains all output produced by `inaccurate_discrimination.do` organized into subfolders for figures, tables, and the log file. The figures folder also includes figures not produced by Stata but used in the paper (`figures/Non-Stata Output`).

Instructions

More complete documentation may be found in the do-files, but brief summaries are provided below:

1. `inaccurate_discrimination.do`: This file converts the raw Qualtrics output (survey responses and payments) to the final analysis dataset. Tables 2, 3, and 4 of the paper are produces in Section 2. Tables B1 – B6 and Figures B1 & B2 of the Online Appendix are produced in Section 3. Additional statistics reported in the paper are produced in Section 4. Tables 2, 3, B4, & B6 may be produced after running the do file by typing in Stata: `texdoc do "${do}inaccurate_discrimination.do"` (after first installing the texdoc package by typing: "`ssc install texdoc`"). The version of Tables B1 and B5 used in the Appendix (`tableB1_final.tex` & `tableB5_final.tex`) require some additional hand formatting to produce from the raw files (`tableB1.tex`, `tableB5_sumstats.tex`, & `tableB5_diff.tex`)

2. `1-Process-Part-1.R`: This file lightly processes the results from Survey 1 and produces the profile data used tor produce the images shown to employers in Survey 2 (`Profiles.csv`).

3. `Generate-Profiles.R`: This file produces the 589 images (PNG files; located in "/Data-Extract/Profile-Pictures") sampled and displayed in Survey 2.

Software and hardware requirements

- Stata (code was last run with version 17.1).

- The following Stata packages may be installed by typing "`ssc install`" followed by the package name below:
  - `estout`
  - `winsor2`
  - `texdoc`

- Stata do files were run on a Macbook Pro laptop with a 2.4 GHz 8-Core Intel Core i9 processor and 64 GB 2667 MHz DDR4 memory. However, these files can be processed with much lower computing power & memory (final dataset is 14.6 MB).

Contents

File Name	Description	Provided
	Cleaning & Analysis Files (Located in Scripts Directory)	
inaccurate_discrimination.do	This is the primary cleaning and analysis file that produces all the tables, figures, and statistics reported in the paper and appendix.	Yes
1-Process-Part-1.R	This file lightly processes the results from Survey 1 and produces the profiles shown to employers in Survey 2	Yes
Generate-Profiles.R	Creates the profile images from Profiles.csv	Yes
	Raw Data Files (Located in Data-Raw Directory)	
Qualtrics-Part1.csv	Survey response output from Qualtrics for Survey 1 (Workers)	Yes
Qualtrics-Part2.csv	Survey response output from Qualtrics for Survey 2 (Employers)	Yes
MTurk-India-Part1.csv	MTurk payments for India sample in Survey 1	Yes
MTurk-US-Part1.csv	MTurk payments for US sample in Survey 1	Yes
MTurk-India-Part2.csv	MTurk payments for India sample in Survey 2	Yes
MTurk-US-Part2.csv	MTurk payments for US sample in Survey 2	Yes
LitSurveyData.pdf	Bibliography for all papers included in the broad literature review	Yes
	Analysis Data Files (Located in Data-Extract Directory)	
inaccurate_discrimination.dta	This is the final cleaned analysis dataset corresponding to the sample used in all analysis.	Yes
Profiles.csv	This file is produced by 1-Process-Part-1.R and includes just the output from Survey 1 formatted to be used in the worker profiles sampled and shown to employers in Survey 2.	Yes
cleaned_worker.dta	Intermediate file created in cleaning process	Yes
cleaned_employer.dta	Intermediate file created in cleaning process	Yes
mturk_worker_US.dta	Intermediate file created in cleaning process	Yes
mturk_worker_India.dta	Intermediate file created in cleaning process	Yes
mturk_worker.dta	Intermediate file created in cleaning process	Yes
mturk_employer_US.dta	Intermediate file created in cleaning process	Yes
mturk_employer_India.dta	Intermediate file created in cleaning process	Yes
mturk_employer.dta	Intermediate file created in cleaning process	Yes