When Is Discrimination Unfair

Replication Package

1_Data Construction

This document provides an overview of the steps taken to convert the raw survey data set into a usable format for conducting the analyses of the paper. It provides a description of the relevant files needed to organize the raw data set and a list of steps to conduct this task.

Description of Files

This folder contains the following files:

aa_when_is_discrimination_unfair_raw_data.csv *(required)*

> File **aa** is the raw data set for the paper – i.e., it is output from Qualtrics, the program used to administer the survey.

bb_when_is_discrimination_unfair_organize_data.do *(required)*

> File **bb** is a do-file that organizes **aa** into a usable format for conducting the paper's analyses – its final output is file **d1_when_is_discrimination_unfair_data_set.dta** in folder 2_Main Analyses.

cc_acs_weights.dta *(optional)*

> File **cc** is a dta file that contains the shares of individuals from the 2019 ACS of certain (age x education x race x gender) cells. This file is used to calculate weights based on this survey.
>
> It is the output from R file **jj.vii**.

dd_first_stage.csv *(required)*

> File **dd** is a csv file that indicates whether each observation (i.e., a scenario from each respondent) from the organized data set belongs in Stage 1 or 2 of the survey.

ee_gss_weights.dta *(optional)*

> File **ee** contains survey weights based on the 2020 GSS.
>
> It is the output from R file **jj.viii**.

ff_sample_weights.dta *(optional)*

> File **ff** contains the "sample" shares of individuals from the MTurk sample of certain (age x education x race x gender) cells. It is also used to develop survey weights based on the 2019 ACS.
>
> It is the output from R file **jj.ix**.

gg_state_data.dta *(required)*

File **gg** contains state and census region data for each MTurk respondent.

hh_suspicious_variables2.dta *(required)*

File **hh** flags all respondents we suspect were automated or extremely low quality.

ii_mturk_data_for_weights.csv *(optional)*

File **ii** indicates the (age x education x race x gender) cell for each MTurk respondent.

It is the output from Do file **bb**.

jj_Materials for weights *(required)*

File **jj** is a folder containing material for developing the survey weights. Within this folder are a set of files.

i. acs2019.dta *(required)*
[source: CEPR Data – ACS, "cepr_acs_2019.zip," https://ceprdata.org/acs-uniform-data-extracts/acs-data/]
File **jj.i** is the 2019 ACS sample provided by CEPR.

ii. gss2020.dta *(required)*
[source: General Social Survey, "GSS 2016-2020 Panel," https://gss.norc.org/get-the-data/stata]
File **jj.ii** is the 2020 GSS sample (derived to only the key variables needed for the weights).

iii. mturk_data_for_weights.csv *(required before running file **jj.ix**)*
File **jj.iii** is the same as file **ii**. It is the output from Do file **bb**.

iv. surveys_organize_for_weights.do *(required)*
File **jj.iv** is a do-file that augments **jj.i** and **jj.ii** into a similar format as **jj.iii**. The outputs are **jj.v** and **jj.vi**.

v. asc_data.csv *(optional)*
File **jj.v** is the outputs of do file **jj.iv**.

vi. gss_data.csv *(optional)*
File **jj.vi** is the outputs of do file **jj.iv**.

vii. acs_weights.R *(required)*
Files **jj.vii** takes **jj.v** as input to calculate the "population" shares of ACS respondents across demographic and political leaning cells. The output is **cc.**

viii. gss_weights.R *(required)*
Files **jj.viii** takes **jj.vi** as input to calculate the "population" shares of GSS respondents across demographic and political leaning cells. The output is **ee**.

ix. sample_weights.R *(required)*
Similarly, file **jj.ix** takes **jj.iii** as an input to calculate the "sample" shares of respondents within cells of demographic characteristics. The output is **ff**.

kk_census_state_pop2019.csv *(required)*

File **kk** is a csv file that contains state population from U.S. Census Bureau.

[source: U.S. Census Bureau, https://www.census.gov/data/tables/time-series/demo/popest/2010s-state-detail.html under "Population Estimates by Age (18+): July 1, 2019"]

ll_when_is_discrimination_unfair_prep_state_pop.do *(required)*

File **ll** is a do-file that organizes **kk** into a usable format for conducting the paper's analyses – its output is file **d7_state_pop.dta** in 2_Main Analyses folder.

How to prepare the Data Set from its Raw Version

1. Save files **aa**, **bb**, **dd**, **gg**, and **hh** to some folder. Similarly, save files **jj.i-jj.ix** to some folder. Also, save files **kk** and **ll** to some folder. They may all be in the same location if you wish.

2. Open file **bb** and make sure that the directory is set properly (i.e, check the "cd" command). The do-file must access **aa, dd, gg**, and **hh**.

3. Check line 329 since it merges in data from file **gg** using the "merge" command.

4. Run lines 1-513 of do-file **bb** – i.e., through the chunk of code commented as "Export a data set used to create weights." This last chunk of code spits out file **ii (jj.iii)**. Do not close out of file **bb** yet.

5. (Optional) Run do file **jj.iv** to create files **jj.v** and **jj.vi**. Make sure the directory within the do file is set appropriately to access the ACS and GSS data sets (**jj.i** and **jj.ii**).

6. (Optional) Run the R code within **jj.vii-jj.ix**. Make sure the directory within these R files are set appropriately so they can access access **jj.v**, **jj.vi**, and **jj.iii**, respectively. They provide **cc**, **ee**, and **ff** in csv format.

7. (Optional) Save the ouputs **cc**, **ee**, and **ff** dta files and name them appropriately if necessary (e.g., "cc_acs_weights.dta"). Make sure these are stored in locations that could be accessed by do file **bb**.

8. Run the remaining lines of do file **bb**. These merge files **cc**, **ee**, and **ff**, so make sure the directories are set appropriately within the "merge" commands (look at lines 516-544).

9. Run do file **ll** to create **d7_state_pop.dta** in 2_Main Analyses folder.

Steps 5-7 demonstrate how the survey weights were developed, and they are optional. If you save files **cc, ee, ff** within the same folder as **bb**, you don't need to do step 5-7, and you can just run the entire file **bb** to obtain the organized data set.