

Interactive Intelligence

Checklist for Review of Dataset

(Version 1)

We recommend that students or employees wishing to publish on their data and results for a given research project in the form of a dataset asks a fellow student or colleague to review this dataset with regard to the points in this checklist. The purpose of the checklist is to ensure that all data that can be made available is made available, that all analyses were conducted conscientiously by the researchers, that all results are reported accurately, and that all methods are transparent and sufficiently clear to be reproducible.

If you choose to have your code reviewed according to this checklist, we advise you to upload this document together with your dataset to the research data repository of your choice (e.g. 4TU Research Data) upon publication of your work.

I. Basic Data

| | |
|--|------------------------------|
| Paper title: | Motivating, your way. |
| Name(s) of researcher(s): | Ramya Ghantasala |
| Name of the reviewer: | Alkis Antoniadis |
| Data repository platform (e.g. 4TU Centre for Research Data): | 4TU Centre for Research Data |

II. Checklist

| Statement | Yes | No |
|--|---|--|
| 1. The dataset contains a README file that fulfils the requirements of the data repository platform that the researcher wishes to use. If no such requirements can be found, the dataset nonetheless contains a README file that clearly explains the contents of the dataset? | README mostly meets requirements | https://data.4tu.nl/info/fileadmin/user_upload/Documenten/Guidelines_for_creating_a_README_file.pdf Missing 1.6, 2.1, and 4 (not sure if applicable). |
| 2. Either within the README file or within an extra, easily findable file, the researchers have explained their data. This means that, for example, for every column of a tabular dataset, all column names and possible cell values are explained. | .xlsx file describes dataset columns. | |
| 3. data is in readily readable file formats. If this should not be the case, the README (or similar) clearly explains the file format and which software can be used to access the contents. | .csv file format is accessible. | |
| 4. All data has been anonymized in accordance to promises made in the Data Management Plan. | Do not know what promises were made in the DMP, but the data is anonymized. | |
| 5. The analysis file or files contain a header with meta-data (name of author, date of writing, required input files and generated output files). | Name of author and date. | No input or output files mentioned. |
| 6. All required input files for the analysis are available in the dataset. | Analysis data files present. | |

| Statement | Yes | No |
|--|---|--|
| 7. There is an output file that is generated by the analysis script that neatly combines code and commentary (e.g. markdown output file). This output file is in a readily readable file format (e.g. pdf). | If this refers to the knitted files, then technically no, it's not provided (assuming that the ones provided are only for me to compare the values), but instructions are there to generate it. | |
| 8. The analysis script is clean and comprehensible in the sense that: <ul style="list-style-type: none"> There is sufficient, useful, and clearly written commentary Irrelevant code (such as old analyses) has been removed The details of analyses that are not reported in the paper (e.g. assumption checks) are proportional to those that are reported in the paper. This means that unreported analyses should not clutter up the script, making it long and unreadable. | Analysis is generally explained. | 1. No code shown in knitted pdf for how posterior probabilities are calculated (only result is shown). 2. In "Statistical Analysis.Rmd" file, WAIC plot is not displayed correctly. 3. In the "Demographic Analysis.Rmd" file, not entirely clear what the table reports at the "Physical Activity Stage" part. 4. No explicit tie mentioned between code or data, and the two png images provided. |
| 9. The analysis script can be run successfully. | They run from rstudio. | Render command given does not work. Error in <code>setwd("~/analysis")</code> : cannot change working directory |
| 10. All preprocessing steps are clearly described and traceable, especially when preprocessing code cannot be executed because raw data is not available. | Preprocessing script mentioned by name. | 1. Nothing mentioned about the process. 2. No instructions to replicate this step. |
| 11. The analyses and results reported in the manuscript can be found back in the analysis script with labels according to where they appear in the manuscript. | Results appear as outputted by the analysis scripts and are referenced. | |
| 12. All results reported in the manuscript accurately correspond to the output produced by the analysis script. | "Prior Sensitivity Analysis.Rmd" and "Demographic Analysis.Rmd" results correspond to those reported in the manuscript. | "Statistical Analysis.Rmd" results are inconsistent with what is reported in the manuscript. |

III. Additional comments by reviewer

Please state any additional things you noticed in reviewing the dataset or possible points of improvement for the reviewer.

1. I have run the code multiple times, even deleting the docker image and building a new one, and the precis function in the "Statistical Analysis.Rmd" file produces different results than the ones in the knitted file provided. I get the same results every time, and they are not the same as the ones reported/provided.
2. It would be a good idea to have instructions for alternative methods that one can use to run the code and knit the files.

IV. Review log

| Round | Date | Paper Status | Checklist Items | Signature Reviewer | Signature Researcher |
|-------|----------|---|-----------------|--------------------|----------------------|
| 1 | 07/03/22 | Draft MSc thesis sent to thesis committee for further feedback. | 12 | A.Antoniades | |
| | | | | | |
| | | | | | |
| | | | | | |
| | | | | | |