

High-dimensional-mediation-analysis

High-dimensional mediation analysis for selecting DNA methylation Loci mediating childhood trauma and cortisol stress reactivity

Supplementary material to the manuscript *High-dimensional mediation analysis for selecting DNA methylation Loci mediating childhood trauma and cortisol stress reactivity*.

1. Download data and set up correct paths

- Step 1. Download data E-GEOD-77445.processed.1.zip (<https://www.ebi.ac.uk/arrayexpress/files/E-GEOD-77445/E-GEOD-77445.processed.1.zip>) as well as E-GEOD-77445.processed.2.zip (<https://www.ebi.ac.uk/arrayexpress/files/E-GEOD-77445/E-GEOD-77445.processed.2.zip>). Unzip these two zip files in the data folder.

Note that there is no need to download `E-GEOD-77445.sdrf.txt`, which has different row length and cannot be read into R directly. We have downloaded and manually cleaned it. We saved it as `X.tsv` in the `data` folder because this file contains information about exposure variable.

- Step 2. Open High-dimensional-mediation-analysis-JASA.Rproj. This step will help you to correctly setup working directory.

Now you are ready to run the data analysis related to case study in Section 3

2. Real Data Analysis

- Step 1. Open case_study_main.Rmd in the Rstudio.

Before running this R markdown, you have to set the `data` folder's absolute path correctly.

To this end, we may have to change the path

```
PATH = "E:/JASAACS20210115Code/data/"
```

in the R markdown file to the corresponding one in which you store your files.

Caution: the PATH should end with `/` or `\\`.

- Step 2. After setting the path correctly, you can click Knit (in the upper left corner) to run the entire markdown.

Alternatively, you may run the R code in individual cells one by one.

You may need to wait for a few minutes to knit this R markdown. For your convenience, we have saved its output as case_study_main.html and case_study_main.pdf. Thus, you may have a quick check.

3. Run simulation for this paper

- Step 1. Open High-dimensional-mediation-analysis-JASA.Rproj. This step will help you to correctly setup working directory.
- Step 2. Open and run simulationPreprocessing.R in R studio
- Step 3. Open and run Simulation_no_confounders.R for simulation studies without confounding variables (section 4.1) This may take two weeks to finish the simulation in this part.
- Step 4. Open and run Simulation_confounders.R for simulation studies with confounding variables (section 4.2) This may take two weeks to finish the simulation in this part.

- Step 5. Open and run `Simulation_compare_sigma1.R` for comparing $\hat{\sigma}_1$ by three methods: our proposed new method, oracle and Zhou et al (2020) (section 4.2 Figure 3) This may take 10 hours to finish the simulation in this part.
- Step 6. Open and run `Simulation_global_test_compare.R` for comparing our method with oracle model, Zhou et al. (2020) and global test for both non-confounding and confounding cases (section S.3) This may take two months to finish the simulation in this part.

R code in this folder are:

1. `utils_mediation.R` stores all the functions that will be used in `case_study_main.Rmd`.
2. `simulationPreprocessing.R` loads the processed data and prepare for the simulations (section 4). Please run through this file before the following files, because all the simulations depend on this file's result.
3. `Simulation_no_confounders.R` performs simulation studies without confounding variables (section 4.1).
4. `Simulation_confounders.R` performs simulation studies with confounding variables (section 4.2).
5. `Simulation_compare_sigma1.R` compares estimated $\hat{\sigma}_1$ among three methods: our proposed new method, oracle and Zhou et al (2020) (section 4.2 Figure 3).
6. `Simulation_global_test_compare.R` compares our method with oracle model, Zhou et al. (2020) and global test for both non-confounding and confounding cases (section S.3).

Other files:

7. 'case_study_main.Rmd' is a R markdown file for real data analysis
8. `case_study_main.html` is a markdown file generated by `case_study_main.Rmd` which is easy-to-read in any browser (such as Chrome).
9. `case_study_main.pdf` is a markdown file generated by `case_study_main.Rmd`.
10. `README.md` is a markdown file providing the instruction of using R markdown and R code in this folder.
11. `README.html` is a file generated by `README.md`
12. `README.pdf` is a file generated by `README.md`

Note that `simulationPreprocessing.R` file requires R package `freebird`, which should be installed from the GitHub. If you fail to install `freebird` package, this likely is because your computer has an old version of `Rtools`. To update `Rtools`, you may go to the following link:

<https://cran.r-project.org/bin/windows/Rtools/rtools40.html>

and follow its instruction to install `Rtools` and setup it correctly.

- After installing and setting path of `Rtools` successfully, rerun codes in line 6-13 of `simulationPreprocessing.R`.