AMS 597: Statistical Computing

Pei-Fen Kuan (c)

Applied Math and Stats, Stony Brook University

- Each student will write an R package implementing statistical methods described in the following paper "A Simple and Robust Method for Partially Matched Samples Using the P-Values Pooling Approach", Stat Med (2013)
 - https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3717400/
- The pdf version of this paper is available at Blackboard (see ScientificPaper_ProjectSpring2021.pdf)

- This paper discussed several methods for handling partially matched samples
- E.g., Complete samples

```
Before Trt
                 x_1
                        x_2
                               x_3
                                     x_4
                                            x_5
                                                   x_6
                                                         x_7
                                                                x_8
                                                                       x_9
                                                                             x_{10}
After Trt
                 y_1
                        y_2
                               y_3
                                     y_4
                                            y_5
                                                   y_6
                                                         y_7
                                                                y_8
                                                                       y_9
                                                                             y_{10}
```

• Partially matched samples

```
Before Trt
                            NA
                                   NA
                                          NA
              x_1
                     x_2
                                                 x_6
                                                      x_7
                                                            x_8
                                                                       x_{10}
                                                                  x_9
After Trt
              NA
                     NA
                            y_3
                                   y_4
                                          y_5
                                                 y_6
                                                       y_7
                                                            y_8
                                                                  y_9
                                                                       y_{10}
```

- You will write function(s) implementing
 - ► Liptak's weighted Z-test
 - Kim et al.'s modified t-statistic
 - ▶ Looney and Jones's corrected Z-test
 - ▶ Lin and Stivers's MLE based test under heteroscedasticity
 - Ekbohm's MLE-based test under homoscedasticity
- You will then wrap these up as an R package

- The R package has to be complete and contains a vignette describing how to use the R package
- The R package is due May 05, 2021 at 5:00 PM
- Submit your package as original source package (i.e., .tar.gz file) on Blackboard>Assignments>Project. Name your package PMlastname_version.tar.gz (version is generated automatically after you build your package successfully)
- Some of the grading criteria include:
 - ► Can the R package be installed successfully?
 - ▶ Is the R package implementing the required method correctly?
 - ► Has it considered all possible scenarios?
 - ▶ Is the R package user friendly (vignette, help files, warning messages, sample data, sample code)?
 - ▶ What is the computational speed?

- Some useful links:
- http://kbroman.org/pkg_primer/
- http://kbroman.org/Tools4RR/assets/lectures/08_rpack_ withnotes.pdf
- https://hilaryparker.com/2014/04/29/ writing-an-r-package-from-scratch/
- https://cran.r-project.org/doc/contrib/ Leisch-CreatingPackages.pdf
- https: //ourcodingclub.github.io/tutorials/writing-r-package/