benchtestr

Igor Geyn and Shing Hon Lam

6/1/2021

What did we do?

- Collect datasets for benchmarking experimental findings
- Write up a variety of classical estimators (DIM, linear regression, matching, etc.)
- Draft up instructional documentation (vigneettes)
- **Compile everything into an R package called `benchtestr`**

Motivation

- Experiments are common throughout (social) science disciplines, and proliferating
- Past work has shown miscalibration in observational findings using experimental benchmarks
  - LaLonde (1986); Dehejia and Wahba (1999)
  - Green, et al (2009)
  - Gerber and Green (2000); Imai (2005)

What is `benchtestr`?

- One-stop shop for benchmarking experimental findings (test new estimators, compare across datasets, etc.)
- Robust support for matching (balance testing, estimation, etc.)
- Visualization (esp. multi-outcomes analyis)
- Saving time (let the defaults do the work)

Examples

- National Supported Work Demonstration (NSW) based on Dehejia and Wahba, LaLonde
- Tennessee Student Teacher Achievement Ratio (STAR) program
- Let's take a look (vignette)

What else could we do?

- Lots of things:
  - Continue to add estimators and data
  - Build out reporting capabiities (push 'go' for a single summary)
  - Build out capacity for sensitivity analysis
  - Push for CRAN-readiness