

Survey Analysis in Julia

Introduction to Survey.jl

Shikhar Mishra

xKDR Forum

22 March 2023

Motivations

What is complex survey analysis?

- Surveys are an empirical tool for social, behavioural and experimental sciences
- Goal: obtaining estimates for a population by "surveying" a well selected sample
- Special techniques available for increasing precision and representation of the surveyed sample
 - ▶ several types of survey "designs" and sampling schemes
- Computing summary statistics from a survey requires applying mathematical corrections and adjustments
 - ▶ eg. population mean is not as simple as arithmetic mean of a numeric vector
- A "survey" package exposes an intuitive API to user, and automatically applies formulae and corrections in background
 - ▶ In Survey.jl, for population mean of a variable you can do `mean(:variable, design)`

Our engineering journey

- Users of R 'survey' package
 - ▶ Benchmark for open-source complex survey analysis
- R 'survey' designed in early 00's for MB's of data
 - ▶ slow for "large" modern datasets and many class of simulation problems
 - ▶ eg. variance estimation using bootstrapping
 - ▶ computation times upto few hours for summary statistics

Why Julia for complex survey analysis

Performance **Verbosity** of R/Python meets **speed** of a systems language

Community Several unmaterialised attempts to create survey analysis package. We received feedback and even contributing PRs on the project. ▶ ¹

Dev & maintenace Avoid "two-language problem". Survey researchers just want something that works great out of the box.

Ecosystem Julia has matured to have substantial statistical computing abilities. Survey.jl is complement to and complemented by the entire data ecosystem.

What our package provides

...

... add more here

Testing against R

Julia 10k sims, R 1k sims in same time or less.

Demo

Getting started with Survey.jl - Tutorial part. Try to atleast add the CPHS Survey.jl code into slides. Do small demo if time/audience.

Future Roadmap

paste some from the homepage of Documentations. Give overview of future issues and PRs github.

Links

- Julia Discourse posts [here](#) and [here](#)
- Unmaterialised attempts [samplics/survey.jl](#) and [jamanrique/SurveyAnalysis.jl](#)