Introducing CAMIS

Comparing Analysis Method Implementations in Software

Chi Zhang, PhD - on behalf of CAMIS WG

UseR! 2024 2024.7.10

About me

Co-lead of CAMIS project since 2024.01

Statistician and part-time lecturer at Faculty of Medicine, University of Oslo

R developer, use Quarto for various things - that's another story!

Today: **CAMIS**

How I joined CAMIS
Non-industry statistician / R dev

Was planning a **career transition**
(academia, public health -> ?)

Have been following updates in the
pharmaceutical industry - **spent a lot of time
on LinkedIn**

September 2023

Harshal (one of the co-leads) invited me to the
project

Not sure about SAS, but I can help with the R
part - and they use **quarto**!

posit   PRODUCTS ⌄   SOLUTIONS ⌄   LEARN & SUPPORT ⌄   EXPLORE MORE ⌄

PHARMA

The future of pharma is open source

Open source is changing how drug development happens. Clinical teams are making better decisions, working through clinical trials more confidently, are bringing life-changing drugs to market faster.

BOOK A CALL WITH OUR PHARMA EXPERTS

CAMIS

Increase understanding and awareness of **analysis result discrepancies** across software (R, SAS, Python etc)

Demonstrate the methodology through **examples**, document in open GitHub repository

Key Considerations When Understanding Differences in Statistical Methodology Implementations Across Programming Languages – An Introduction to the CAMIS Project

Min-Hua Jen, Brian Varney, Kyle Lee, Benjamin Arancibia, Mia Qi, Lyn Taylor, Christina Fillmore, Joseph Rickert, Mike Stackhouse, Michael Rimler

Repository location:
https://psiaims.github.io/CAMIS/

Discrepancies
stata vs R, test for proportions

```
. prtesti 1000 0.123 0.13

One-sample test of proportion                    x: Number of obs =        1000

                    |      Mean    Std. err.              [95% conf. interval]
              ------+-------------------------------------------------------
                  x |      .123    .0103861              .1026436    .1433564

      p = proportion(x)                                         z =   -0.6582
H0: p = 0.13

      Ha: p < 0.13                Ha: p != 0.13                 Ha: p > 0.13
Pr(Z < z) = 0.2552        Pr(|Z| > |z|) = 0.5104        Pr(Z > z) = 0.7448
```

```
> prop.test(x = 123, n = 1000, p = 0.13, alternative = 'less')

        1-sample proportions test with continuity correction

data:  123 out of 1000, null probability 0.13
X-squared = 0.37356, df = 1, p-value = 0.2705
alternative hypothesis: true p is less than 0.13
95 percent confidence interval:
 0.0000000 0.1416364
sample estimates:
    p
0.123
```

```
> prop.test(x = 123, n = 1000, p = 0.13, alternative = 'less', correct = F)

        1-sample proportions test without continuity correction

data:  123 out of 1000, null probability 0.13
X-squared = 0.43324, df = 1, p-value = 0.2552
alternative hypothesis: true p is less than 0.13
95 percent confidence interval:
 0.0000000 0.1411081
sample estimates:
    p
0.123
```

Shift to open source is also happening in **public** sector (if not more!) - $$$

The default **algorithms** under the hood are different

Not well documented in stata

What if the p-values differ **around 0.05**?

Rounding

round(1.5) =

```
> round(1.5)
[1] 2
```

round(0.5) =

```
> round(0.5)
[1] 0
> janitor::round_half_up(0.5)
[1] 1
```

round(1.55) =

```
> round(1.55)
[1] 2
```

round(1.55, digits = 1) =

```
> round(1.55, digits = 1)
[1] 1.6
```

In R, `round()` to the nearest even; in SAS, `round()` round half up

Pharmaverse blog has an article on this topic

	round half up	round to even	round up	round down	round towards zero
Example: 1.45	1.5 (round to 1 decimal place)	1.4 (round to 1 decimal place)	2	1	1

Here are the corresponding ways to implement these methods in SAS and R.

	round half up	round to even	round up	round down	round towards zero
SAS	round()	rounde()	ceil()	floor()	int()
R	janitor::round_half_up() tidytlg::roundSAS()	base::round()	base::ceiling()	base::floor()	base::trunc()

Zhang, Kangjie. 2023. "Rounding." August 22, 2023. https://pharmaverse.github.io/blog/posts/2023-07-24_rounding/rounding.html.

Why it matters

Shift **in industry** from SAS to R and open source software, e.g. pharmaverse

In **public sector**: teaching, medical research and public health, SPSS / stata (even excel!) vs R, python

Do we get the same results? Which is correct, Why do they differ?

Could be different **algorithms**; default **options**; floating point numbers; … (*@adrianolszewski listed at least 14 different reasons*)

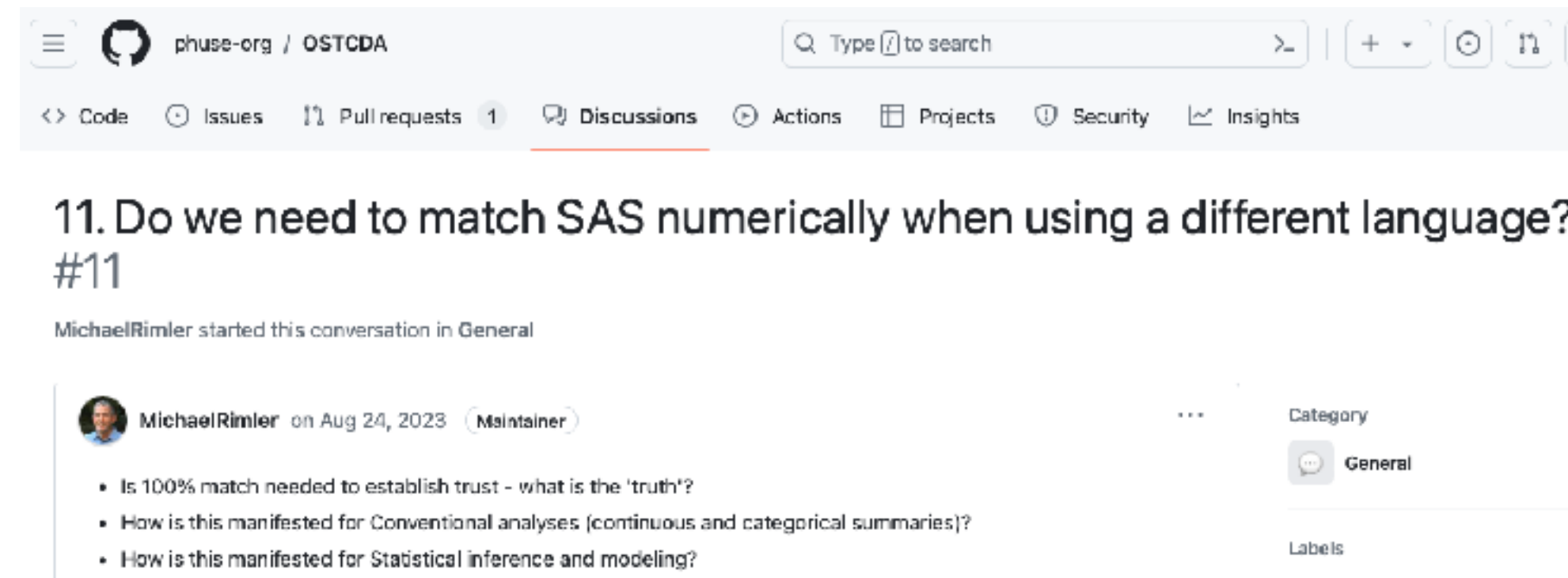When results are different, it leads to **uncertainty**

Reproducibility challenges already exist even when there's *only one software* (set seed, package versioning, inaccessible sensitive data for revision, …)

Trust in open source?

phuse

Open Source
Technology in Clinical
Data Analysis

11. Do we need to match SAS numerically when using a different language?
#11

MichaelRimler started this conversation in General

MichaelRimler on Aug 24, 2023 Maintainer ···

- Is 100% match needed to establish trust - what is the 'truth'?
- How is this manifested for Conventional analyses (continuous and categorical summaries)?
- How is this manifested for Statistical inference and modeling?

Category

💬 General

Labels

Accuracy, reproducibility, traceability (Modernization of Statistical Analytics Framework)

Validation: establish **documented evidence** which provides a high degree of assurance that procedures **consistently** produce a product meeting its predetermined specifications

It doesn't mean we need to match results 100%; but we should know **why** they are different

Rounding
In CAMIS

```
round(1.5) =        > round(1.5)
                    [1] 2


round(0.5) =        > round(0.5)
                    [1] 0


round(1.55) =       > round(1.55)
                    [1] 2


round(1.55, digits = 1) =

                    > round(1.55, digits = 1)
                    [1] 1.6
```

Methods		R SAS Python Comparison
Summary Statistics	Rounding	R SAS Python R vs SAS
	Summary statistics	R SAS Python R vs SAS
	Skewness/Kurtosis	R SAS Python R vs SAS

R v SAS rounding

Rounding; R and SAS

On comparing the documentation of rounding rules for both languages, it will be noted that the default rounding rule (implemented in the respective language's round() function) are different. Numerical differences arise in the knife-edge case where the number being rounded is equidistant between the two possible results. The round() function in SAS will round the number 'away from zero', meaning that 12.5 rounds to the integer 13. The round() function in Base R will round the number 'to even', meaning that 12.5 rounds to the integer 12. SAS does provide the rounde() function which rounds to even and the janitor package in R contains a function round_half_up() that rounds away from zero. In this use case, SAS produces a correct result from its round() function, based on its documentation, as does R. Both are right based on what they say they do, but they produce different results (Rimler, M.S. et al.).

References

Rimler M.S., Rickert J., Jen M-H., Stackhouse M. Understanding differences in statistical methodology implementations across programming languages (2022, Fall). ASA Biopharmaceutical Report Issue 3, Volume 29. Retrieved from https://higherlogicdownload.s3.amazonaws.com/AMSTAT/fa4dd52c-8429-41d0-abdf-0011047bfa19/UploadedImages/BIOP%20Report/BioPharm_fall2022FINAL.pdf

Where are we now, where do we go next?

CAMIS today

Contributors 18
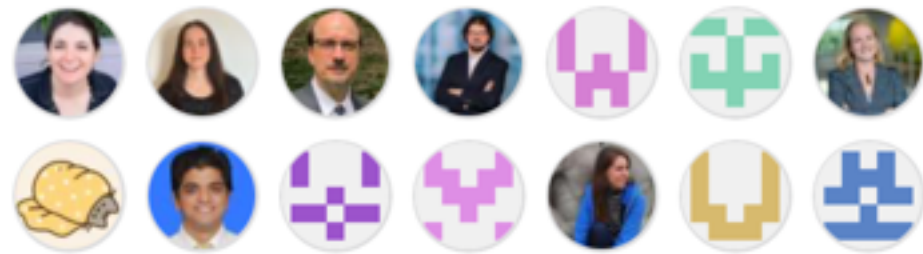
+ 4 contributors

Deployments 187

✅ github-pages 4 days ago

+ 186 deployments

Many participants on the monthly update meetings *(might not directly contribute to the content on GH, but in other ways - discussion, collaboration, presentation)*

Documentation in **R, SAS, Python**

About   Contribute   Publications and projects ▾   News ▾

Motivation

The goal of this project is to demystify conflicting results between software and to help ease the transitions to new languages by providing comparison and comprehensive explanations.

Repository

The repository below provides examples of statistical methodology in different software and languages, along with a comparison of the results obtained and description of any discrepancies.

On this page
Introduction
Motivation
Repository

Methods		R	SAS	Python	Comparison
Summary Statistics	Rounding	R	SAS	Python	R vs SAS
	Summary statistics	R	SAS	Python	R vs SAS
	Skewness/Kurtosis	R	SAS	Python	R vs SAS
General Linear Models	One Sample t-test	R	SAS	Python	R vs SAS
	Paired t-test	R	SAS	Python	R vs SAS
	Two Sample t-test	R	SAS	Python	R vs SAS
	ANOVA	R	SAS		R vs SAS
	ANCOVA	R	SAS	Python	R vs SAS
	MANOVA	R	SAS	Python	R vs SAS
	Linear Regression	R	SAS		R vs SAS
Generalized Linear Models	Logistic Regression	R	SAS		
	Poisson/Negative Binomial Regression	R			

CAMIS today
SAS vs R

Since CAMIS group has a focus on pharmaceutical industry, the **comparisons** are mostly in SAS vs R

Example page: one sample t-test

Agnieszka (our contributor) has an **in-depth analysis** on different implementations on **Wilcoxon Signed Rank Test** - R, StatXact, SAS

Talk recording and slides she gave on *R/ Medicine 2024* is available on CAMIS website

One Sample t-test Comparison

The following table shows the types of One Sample t-test analysis, the capabilities of each language, and whether or not the results from each language match.

Analysis	Supported in R	Supported in SAS	Results Match	Notes
One sample t-test, normal data	Yes	Yes	Yes	In Base R, use `mu` parameter on `t.test()` function to set null hypothesis value
One sample t-test, lognormal data	Maybe	Yes	NA	May be supported by **envstats** package

Comparison Results

Normal Data

Here is a table of comparison values between `t.test()`, `proc_ttest()`, and SAS `PROC TTEST`:

Statistic	t.test()	proc_ttest()	PROC TTEST	Match	Notes
Degrees of Freedom	29	29	29	Yes	
t value	2.364306	2.364306	2.364306	Yes	
p value	0.0249741	0.0249741	0.0249741	Yes	

CAMIS today

About

🔗 psiaims.github.io/CAMIS/

📖 Readme

⚖️ Apache-2.0 license

∿ Activity

▣ Custom properties

☆ 43 stars

👁 6 watching

⑂ 21 forks

Report repository

Closed 37 issues

187 deployments of the website

Still many open issues - we need your help!

☐ ◉ 46 Open ✓ 37 Closed   Author ▾

☐ ◉ **Pearson's/ Spearman's/ Kendall's Rank** `Python`
#216 opened 3 weeks ago by seemaniabhilipsa

☐ ◉ **Website: dissertation info** `documentation`
#215 opened 3 weeks ago by andreaczhang ◐ 2 of 4 tasks

☐ ◉ **MANOVA** `Python`
#208 opened on May 21 by seemaniabhilipsa

☐ ◉ **Overall structure**
#200 opened on May 6 by therneau

☐ ◉ **Survey Statistics - Example/Comparison (Python)** `Python`
#185 opened on Apr 15 by michaelwalshe

☐ ◉ **add more contribution guidance**
#143 opened on Feb 7 by DrLynTaylor

☐ ◉ **Add more information about the version of R running and package**
#139 opened on Jan 25 by statasaurus

☐ ◉ **Other Methods - Machine learning** `Comparison` `R` `SAS`
#67 opened on Mar 20, 2023 by statasaurus

☐ ◉ **Other Methods - Causal inference** `Comparison` `R` `SAS`

We use Quarto

Quarto has been the **publishing and collaboration** tool we use

Easy to write articles, render results for multiple language (R, python)

For SAS it is a mix of **copy-paste** and screenshots

Website deployment is straightforward when only text is edited

R packages and versioning: we're working on a solution with **renv**, getting help from posit!

```
proc freq data = test_case;
weight Count;
tables treatment * Weight / chisq fisher;
exact or;
run;
```

Output:

```
proc means data=htwt;
run;

                Descriptive Statistics for HTWT Data Set
                        The MEANS Procedure

Variable  Label     N        Mean        Std Dev       Minimum       Maximum
-------------------------------------------------------------------------------
AGE       AGE       237     16.4430380     1.8425767    13.9000000     25.0000000
HEIGHT    HEIGHT    237     61.3645570     3.9454019    50.5000000     72.0000000
WEIGHT    WEIGHT    237    101.3080169    19.4406980    50.5000000    171.5000000
-------------------------------------------------------------------------------
```

Statistics for Table of treatment by Weight

Statistic	DF	Value	Prob
Chi-Square	1	2.3072	0.1288
Likelihood Ratio Chi-Square	1	2.2490	0.1337
Continuity Adj. Chi-Square	1	1.8261	0.1766
Mantel-Haenszel Chi-Square	1	2.2964	0.1297
Phi Coefficient		0.1041	
Contingency Coefficient		0.1035	
Cramer's V		0.1041	

Active and welcoming community

4 co-leads: **Lyn Taylor** (Parexel), **Christina Fillmore** (GSK), **Harshal Khanolkar** (Novo Nordisk), **Chi Zhang** (University of Oslo)

Community call once per month: discuss open issues, updates

Open and organized: meeting minutes openly accessible

Diverse and great vibes!

We encourage everyone to go out and present our group at conferences and seminars

Active and welcoming community

4 co-leads: **Lyn Taylor** (Parexel), **Christina Fillmore** (GSK), **Harshal Khanolkar** (Novo Nordisk), **Chi Zhang** (University of Oslo)

Community call once per month: discuss open issues, updates

Open and organized: meeting minutes openly accessible

Diverse and great vibes!

We encourage everyone to go out and present our group at conferences and seminars

Conference name	Date (2024)	Location	Name Attending	Details	Website
RSS Local Group Seminar	28 Feb	Sheffield, England	Lyn Taylor	Slides	RSS
phuse US Connect	25-28 Feb	Bethesa, Maryland, USA	Soma Sekhar Sriadibhatla, Vikash Jain, Brian Varney	Poster	Connect
phuse chapter connect	03 APR	Bangalore	Harshal Khanolkar		
phuse/FDA CSS	3-5 June	Silver Spring Maryland, USA	Mike Stackhouse	CAMIS Discussion	CSS
R/Medicine	10-14 June	Online	Agnieszka Tomczyk, Lyn Taylor	Part1 and Part2 and slides	R/Medicine 2024
UseR!	8-11 July	Salzburg, Austria	Chi Zhang	Presentation	useR! 2024
phuse EU	11-13 Nov	Strasbourg, France	Agnieszka Tomczyk, Christina Fillmore	Presentation	PHUSE EU Connect

Ongoing discussions

Philip Bowsher · 1st
Director, Health and Life Sciences Industry Leader at Posit/RStudio PBC Ta...
3w · Edited · 🌐

Last week I attended a large pharma statistical programming conference (PharmaSUG) and a popular topic was CAMIS for understanding the source of differences between statistical software! Below is an overview:

CAMIS Website: https://lnkd.in/gpRNGFCg

In 2015, the FDA released the Statistical Software Clarifying Statement (https://lnkd.in/gJug8S2u) which states that the FDA does not require use of any specific software for statistical analyses.

Many organizations (Roche, Novo Nordisk, GSK etc.) are keen to use open source for Clinical Study Reports and to generate TFLs in the regulatory space.

One challenge has been observing differences across languages especially for complex TFLs.

Enter "Comparing Analysis Method Implementations in Software" CAMIS!!

The goal of CAMIS is to help understand the source of any differences between software and provide comparison and comprehensive explanations.

The team has collected and documented many SAS & R differences such as:

Repeated Measures Analysis:
https://lnkd.in/ggfTFqjP

Rounding:
https://lnkd.in/gZtFZrps

You and 438 others   23 comments · 38 reposts

**Stephen Senn** · 2nd   3w ···
Statistical Consultant

What a fascinating and important initiative. A classic paper is this by Reinhard Bergmann et all https://www.tandfonline.com/doi/abs/10.1080/00031305.2000.10474513 . Another issue that surfaces is that different packages can use different approaches to parameterisation, especially in the presence of interactions. This can mean that they disagree in estimation even though they agree in prediction. The long standing debate on Type III v Type II SS is related to this. See discussion in https://journals.sagepub.com/doi/10.1177/009286150003400222 of an example of Christy Chiung-Stein's and Donald Tong.

THE AMERICAN
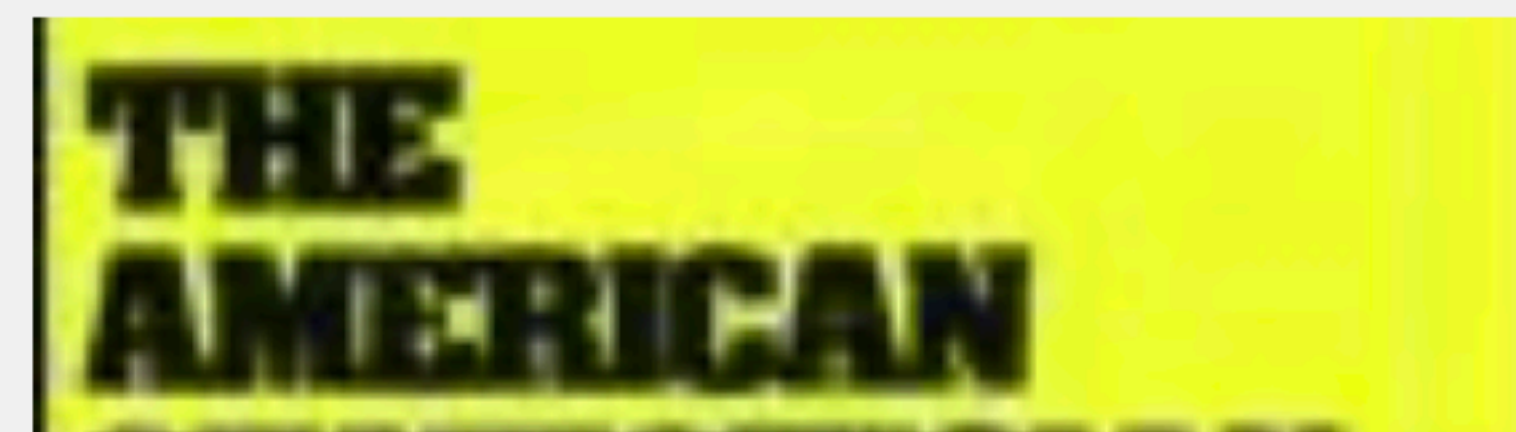
Different Outcomes of the Wilcoxon—Mann—Whitney Test from Different St...
tandfonline.com

Like · 👍⏱ 13 | Reply · 4 Replies

Full discussion and post:

Ongoing discussions

PHUSE Open Source Technology in Clinical Data Analysis WG

https://github.com/phuse-org/OSTCDA/discussions/

They have a community call to discuss relevant topics

phuse

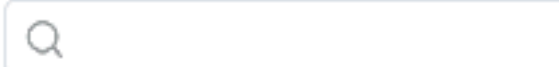Open Source Technology in Clinical Data Analysis ⬇ ⤬

🔍

4 Documenting Trust

4.1 How do you document your trust in an open source solution?

- How do we have document our trust that an open source solution is accurate?

- How do we know if a third-party will accept our documentation of trust?

10 Numerical Matching

10.1 Do we need to match SAS numerically when using a different language?

- What if we the same inputs yield similar, but numerically different results?

- What if we the same inputs yield drastically different results?

- What is the truth? Which is correct?

- What if SAS and R are equivalent, but a third language yields numerical differences?

10.2 How to Contribute

Contribute to the discussion here in GitHub Discussions:
Do we need to match SAS numerically when using a different language?

10.3 Guidance

- Provide your thoughts and perspectives

- Provide references to articles, webinars, presentations (citations, links)

- Be respectful in this community

Collaboration
openstatsware - mmrm

R vs SAS MMRM

Introduction

In this vignette we briefly compare the `mmrm::mmrm`, SAS's `PROC GLIMMIX`, `nlme::gls`, `lme4::lmer`, and `glmmTMB::glmmTMB` functions for fitting mixed models for repeated measures (MMRMs). A primary difference in these implementations lies in the covariance structures that are supported "out of the box". In particular, `PROC GLIMMIX` and `mmrm` are the only procedures which provide support for many of the most common MMRM covariance structures. Most covariance structures can be implemented in `gls`, though users are required to define them manually. `lmer` and `glmmTMB` are more limited. We find that `mmrm` converges more quickly than other R implementations while also producing estimates that are virtually identical to `PROC GLIMMIX`'s.

Convergence Times

FEV Data

The `mmrm`, `PROC GLIMMIX`, `gls`, `lmer`, and `glmmTMB` functions are applied to the FEV dataset 10 times. The convergence times are recorded for each replicate and are reported in the table below.

Comparison of convergence times: milliseconds

Implementation	Median	First Quartile	Third Quartile
mmrm	56.15	55.76	56.30
PROC GLIMMIX	100.00	100.00	100.00
lmer	247.02	245.25	257.46
gls	687.63	683.50	692.45
glmmTMB	715.90	708.70	721.57

It is clear from these results that `mmrm` converges significantly faster than other R functions. Though not demonstrated here, this is generally true regardless of the sample size and covariance structure used. `mmrm` is faster than `PROC GLIMMIX`.

Marginal Treatment Effect Estimates Comparison

We next estimate the marginal mean treatment effects for each visit in the FEV and BCVA datasets using the MMRM fitting procedures. All R implementations' estimates are reported relative to `PROC GLIMMIX`'s estimates. Convergence status is also reported.

FEV Data

Average Treatment Effect Estimates Relative to SAS Estimates
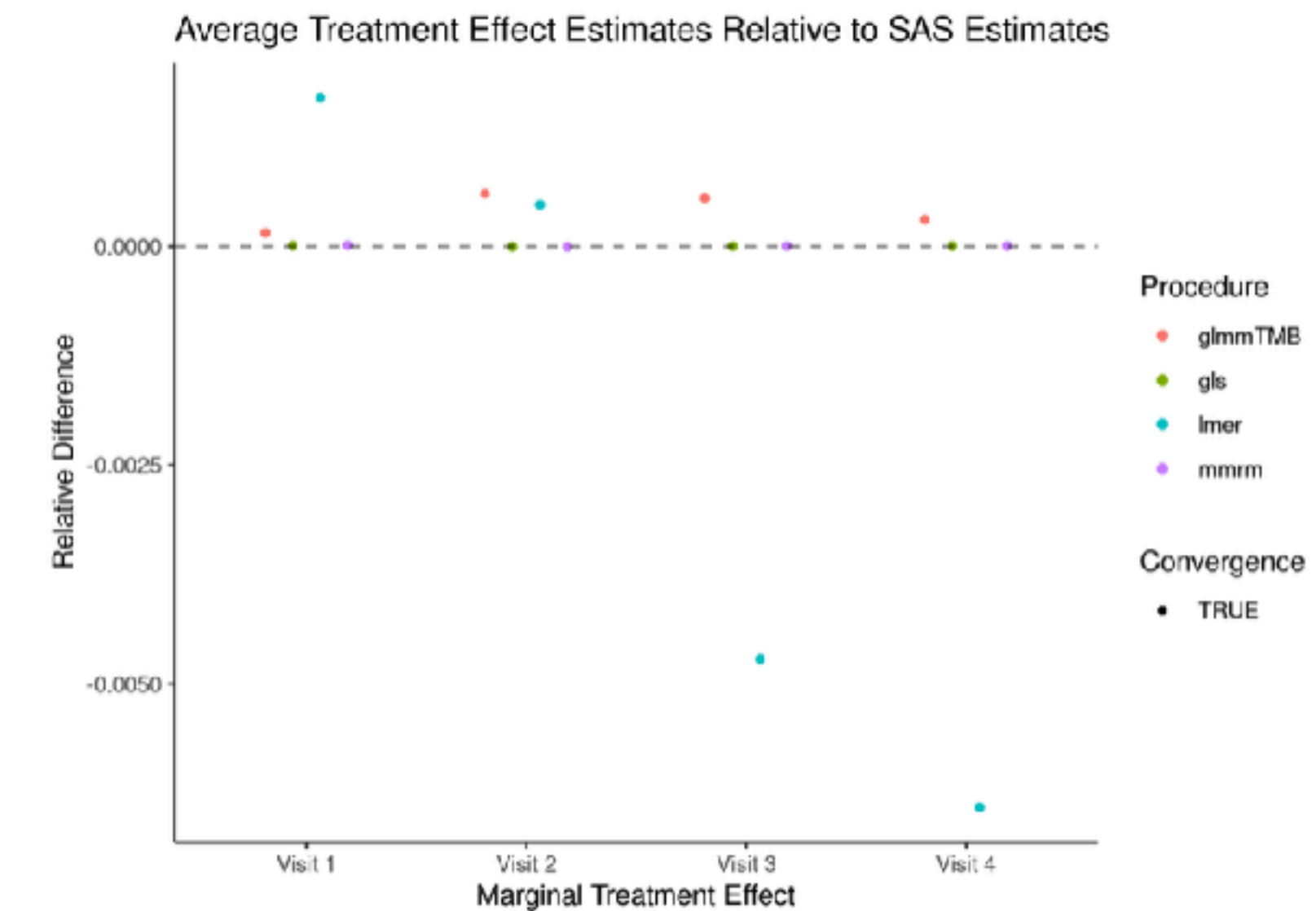
Possibility to add the Bayesian implementation as well

Collaboration
Academia - dissertation

NEW

We are open to **students, researcher** (academia / industry) who want to team up for dissertation projects

Example: *a comparison of MMRM methodology in SAS and R software* (ongoing)

A good combination of **methodology** and **implementation** with real-world **applications**

Contact us for more information!

Repeated Measures	Linear Mixed Model (MMRM)	R SAS	R vs SAS
	Generalized Linear Mixed Model (MMRM)		
	Bayesian MMRM		
Multiple Imputation - Continuous Data MAR	MCMC		
	Linear regression	R	
	Predictive Mean Matching	R	
	Propensity Scores		
Multiple Imputation - Continuous Data MNAR	Delta Adjustment/Tipping Point		
	Reference-Based Imputation/Sequential Methods		
	Reference-Based Imputation/Joint Modelling		
Correlation	Pearson's/ Spearman's/ Kendall's Rank	R	
Survival Models	Kaplan-Meier Log-rank test and Cox-PH	R SAS	R vs SAS
	Accelerated Failure Time		
	Non-proportional hazards methods		
Sample size /Power calculations	Single timepoint analysis		
	Group-sequential designs		
Multivariate methods	Clustering		
	Factor analysis		
	PCA		

CAMIS has something for everyone

❗ Topic of importance for the biopharma industry as well as research

📖 Read CAMIS Documents

😲 Like to be intellectually challenged and fascinated on something subtle

🤝 Collaborate, network and make new friends

🔍 CAMIS

🤓 Learn new topics on statistics and programming

How to get started

Contribution guidelines for new contributors

R template for topics to cover

Also we have information for contributors who are **new to GitHub** and Pull Request

Get started: check our **open issues** and comment to indicate your interest, or send us a message

We aim to close 45 issues this year - 37 so far

Help us achieve the goal :)

How to contribute to the documentation

Please contribute by **submitting a pull request** (PR) and our team will review it.

Adding a new page

If you are adding a new page, please follow our template guideline: R template

Good documentation on data, methods are very much appreciated!

First-time contributors

Welcome to CAMIS! Please read this article: Get started, which contains some useful information to help you navigate your first PR submission.

Asking for help

If you need any assistance with setting up your workspace, do not hesitate to contact @DrLynTaylor, @statasaurus and @andreaczhang!

☐ ⊙ **Sample size /Power calculations - Group-sequential designs** `Comparison` `R` `SAS`
#59 opened on Mar 20, 2023 by statasaurus

☐ ⊙ **Sample size /Power calculations - Single timepoint analysis** `Comparison` `R` `SAS`
#58 opened on Mar 20, 2023 by statasaurus

☐ ⊙ **Survival Models - Non-proportional hazards methods** `Comparison` `R` `SAS`
#57 opened on Mar 20, 2023 by statasaurus

☐ ⊙ **Survival Models - Accelerated Failure Time** `Comparison` `R` `SAS`
#56 opened on Mar 20, 2023 by statasaurus

Resources

CAMIS website: https://psiaims.github.io/CAMIS/

GitHub Repo: https://github.com/PSIAIMS/CAMIS/

Open issues: https://github.com/PSIAIMS/CAMIS/issues

PHUSE discussions

Do we need to match SAS numerically when using a different language?
https://github.com/phuse-org/OSTCDA/discussions/11

How to document your trust in an open source solution?
https://phuse-org.github.io/OSTCDA/doc_trust.html

Please feel free to reach out to the co-leads, you can find us on LinkedIn

Lyn Taylor
Christina Fillmore
Harshal Khanolkar
Chi Zhang

Contact

Website: https://andreaczhang.github.io/
LinkedIn: Chi Zhang
Email: chi.zhang@medisin.uio.no