SESUG Paper 147-2025

Comparing Analysis Method Implementations in Software (CAMIS)

Brian Varney, Experis

ABSTRACT

Several discrepancies have been discovered in statistical analysis results between different programming languages, even in fully qualified statistical computing environments. Subtle differences exist between the fundamental approaches implemented by each language, yielding differences in results which are each correct in their own right. The fact that these differences exist causes unease on the behalf of sponsor companies when submitting them to a regulatory agency, as it is uncertain if the agency will view these differences as problematic. In its Statistical Software Clarifying Statement, the US Food and Drug Administration (FDA) states that it "FDA does not require use of any specific software for statistical analyses" and that "the computer software used for data management and statistical analysis should be reliable." Observing differences across languages can reduce the analyst's confidence in reliability and, by understanding the source of any discrepancies, one can reinstate confidence in reliability.

INTRODUCTION

This paper intends to present the efforts of the CAMIS group. The CAMIS group Comparing Analysis Method Implementations in Software (CAMIS) is a cross-industry PHUSE DVOST Working Group, run in collaboration with members from PHUSE, PSI, ASA and IASCT. In addition to issue comments, which are hosted in the GitHub Repository, we meet monthly on the 2nd Monday of each month. If you would like to join us please contact us at workinggroups@phuse.global.

MOTIVATION

The goal of this project is to demystify conflicting results in statistical analysis methods and results between primarily **SAS**, **R**, and **Python** programming languages by providing comparisons and comprehensive explanations of similarities and differences. Many discrepancies have been discovered in statistical analysis results between these and other programming languages. The differences in results are due to fundamental approaches implemented by each language, which are each correct in their own right. The fact that these differences exist is a challenge, especially related to sponsor companies when submitting them to a regulatory agency.

In its Statistical Software Clarifying Statement, the US Food and Drug Administration (FDA) states that it "FDA does not require use of any specific software for statistical analyses" and that "the computer software used for data management and statistical analysis should be reliable." Observing differences across languages can reduce the analyst's confidence in reliability and, by understanding the source of any discrepancies, one can reinstate confidence in reliability. CAMIS seeks to explore and explain some of the differences and similarities in statical analysis methods between these languages to ease these concerns.

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. Although this is a living and changing repository, the current state is summarized below by topic area.

SUMMARY STATISTICS

- Rounding
- Summary Statistics
- Skewness / Kurtosis

GENERAL LINEAR MODELS

- One Sample T-Test
- Paired T-Test
- Two Sample T-Test
- ANOVA
- ANCOVA
- MANOVA
- Linear Regression

GENERALIZED LINEAR MODELS

- Logistic Regression
- Poisson / Negative Binomial Regression

NON-PARAMETRIC ANALYSIS

- Wilcoxon Signed Rank
- Mann-Whitney U / Wilcoxon Rank Sum
- Kolmogorov-Smirnov Test
- Kruskal-Wallace Test
- Friedman Test
- Jonckheere Test
- Hodges-Lehman Estimator

CATEGORICAL DATA ANALYSIS

- Binomial Test
- McNemar’s Test
- Marginal Homogeneity Tests
- Chi-Square Association / Fishers Exact
- Cochran Mantel Haenszal
- Confidence Intervals for Proportions

REPEATED MEASURES

- Linear Mixed Model (MMRM)
- Linear Mixed Model (degrees of freedom)
- Generalized Linear Mixed Model (GLMM)

- Generalized Estimating Equation (GEE)
- Bayesian MMRM

MULTIPLE IMPUTATION – CONTINUOUS DATA MAR

- MCMC
- Linear Regression
- Predictive Mean Matching

MULTIPLE IMPUTATION CONTINUOUS NMAR

- Tipping Point (Delta Adjustment)
- Reference-Based Imputation / Joint Modelling

CORRELATION

- Pearson's / Spearman's Kendall's Rank.

SURVIVAL MODELS

- Kaplan-Meier Log-Rank Test and Cox PH
- Cause Specific Hazards
- Accelerated Failure Time
- Weighted Log-Rank Test
- Recurrent Events
- Cumulative Incidence Functions
- Tobit Regression
- Restricted Mean Survival Time (RMST)

SAMPLE SIZE / POWER CALCULATIONS

- Intro to Sample Size
- Superiority Single Timepoint
- Equivalence Single Timepoint
- Non-Inferiority Single Timepoint
- Average Bioequivalence
- Cochran-Armitage Test for Trend
- Group Sequential Designs

CAUSAL INFERENCE / MACHINE LEARNING

- Intro to Machine Learning
- Propensity Score Matching
- Propensity Score Weighting
- Clustering
- Factor Analysis
- Principal Components Analysis (PCA)
- Canonical Correlation
- Partial Least Squares (PLS)
- Lasso
- Ridge Regression
- xgboost

OTHER METHODS

- Survey Statistics.

CONCLUSION

Although much has been accomplished with this group, there is much more to do. After learning more about this group and seeing the CAMIS website, we hope that you will sign up to participate and help to fill out our to-do list. How to get Involved:

REFERENCES

- Michael S. Rimler, Joseph Rickert, Min-Hua Jen, Mike Stackhouse. 2022. Understanding differences in statistical methodology implementations across programming languages.
- Statistical Software Clarifying Statement (fda.gov)
- CAMIS White Paper
- How to get involved

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Brian Varney
Experis
269-365-1755
brian.varney@experis.com>