phuse

CAMIS

Comparing

Analysis

Method

Implementations in

Software

phuse

CAMIS

phuse

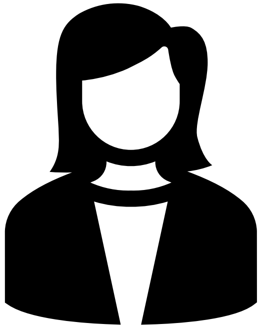
phuse.global

Data Visualisation &
Open Source Technology

Disclaimer

This presentation reflects the view of the PHUSE CAMIS Working Group and does not necessarily represent the position of PAREXEL or GSK

Why Document Comparisons



Lyn
- Has been in industry for 20+ years
- Prefers SAS

Christina
- Newer to industry
- Learned object oriented programming in school and prefers R

Why Document Comparisons

Example 87.4 Hodges-Lehmann Estimation

(View the complete code for this example.)

This example uses the SAS data set *React* created in Example 87.2. The data set con…
variable *Time*, which contains the r…

…hristina

…973), page 34.) Note that in the two-sample case the estimator for the difference in location parameters d…
not estimate the difference in medians (a common misconception) but rather the median of the difference between a sample from `x` and a sample from `y`.

If exact p-values are available, an exact confidence interval is obtained by the algorithm described in Bauer (1972), and the Hodges-Lehmann estimator is employed. Otherwise, the returned confidence interval and poi… estimate are based on normal approximations. These are continuity-corrected for the interval but *not* the esti… (as the correction depends on the `alternative`).

With small samples it may not be possible to achieve very high confidence interval coverages. If this happens… warning will be given and an interval with lower coverage will be substituted.

…hen x (and y if applicable) are valid, the function now always returns, also in the `conf.int = TRUE` ca…

How to Avoid Finding Differences at Critical Times

- Make sure to have **clear requirements**
- Check CAMIS
- Consider defining tolerances rather than looking for exact match between prod and QC

What does CAMIS Cover?

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	Python	R vs SAS
	ANCOVA	R	SAS	Python	R vs SAS
	MANOVA	R	SAS	Python	R vs SAS
	Linear Regression	R	SAS	Python	R vs SAS
Generalized Linear Models	Logistic Regression	R	SAS	Python	R vs SAS
	Poisson/Negative Binomial Regression	R	SAS		R vs SAS
	Categorical Repeated Measures				
	Categorical Multiple Imputation				
Non-parametric Analysis	Wilcoxon signed rank	R	SAS/ StatXact		R vs SAS
	Mann-Whitney U/Wilcoxon rank sum	R	SAS		
	Kolmogorov-Smirnov test				
	Kruskall-Wallis test	R	SAS	Python	R vs SAS
	Friedman test	R	SAS		R vs SAS
	Jonckheere test	R	SAS		R vs SAS
	Hodges-Lehman Estimator	R	SAS		
Categorical Data Analysis	Binomial test	R	SAS	Python	
	McNemar's test	R	SAS		R vs SAS
	Marginal Homogeneity Tests	R			
	Chi-Square Association/Fishers exact	R	SAS	Python	R vs SAS
	Cochran Mantel Haenszel	R	SAS		R vs SAS
	Confidence Intervals for	R	SAS		R vs SAS

		R	SAS	Python	Comparison
Repeated Measures	Linear Mixed Model (MMRM)	R	SAS		R vs SAS
	Linear Mixed Model (degrees of freedom)				
	Generalized Linear Mixed Model (MMRM)				
	Bayesian MMRM				
Multiple Imputation - Continuous Data MAR	MCMC				
	Linear regression	R	SAS		
	Predictive Mean Matching	R			
Multiple Imputation - Continuous Data MNAR	Delta Adjustment/Tipping Point				
	Reference-Based Imputation/Joint Modelling	R	SAS		R vs SAS
Correlation	Pearson's/ Spearman's/ Kendall's Rank	R	SAS	Python	R vs SAS
Survival Models	Kaplan-Meier Log-rank test and Cox-PH	R	SAS		R vs SAS
	Cause Specific Hazards	R	SAS		R vs SAS
	Accelerated Failure Time	R			
	Weighted Log-rank test	R			
	Cumulative Incidence Functions	R	SAS		R vs SAS
	Tobit regression	R	SAS		R vs SAS
	Restricted Mean Survival Time (RMST)		SAS		
Sample size/ Power calculations	Intro to Sample Size				Summary
	Superiority Single timepoint	R	SAS		
	Equivalence Single timepoint	R	SAS		
	Non-Inferiority Single timepoint	R	SAS		
	Average BioEquivalence	R			
	Cochran-Armitage Test For Trend	R	SAS/ StatXact		
	Group sequential designs	R	East		East vs R
Causal inference/ Machine	Intro to Machine Learning				Summary

Differing Defaults

Kaplan Meier

Analysis	Supported in R	Supported in SAS	Results Match	Notes
Kaplan Meier with confidence intervals using log-log method	Yes (using the option conf.type = “log-log”)	Yes (Default)	Mostly	1) Survival estimates can disagree when last event is censored and survival estimate does not cross the percentile being estimated. 2) Survival estimates at time X can disagree when the time X is after the last observed censored time
Kaplan Meier with confidence intervals using log method	Yes (Default)	Yes (using the option conftype=log)	Mostly	As above.

Cox – PH regression

Analysis	Supported in R	Supported in SAS	Results Match
Cox Proportional Hazards Model using breslow method for ties	Yes (using the option ties=“breslow”)	Yes (Default)	Yes
Cox Proportional Hazards Model using efron method for ties	Yes (Default)	Yes (using the option ties=efron)	Yes

Identifying Common Coding Mistakes

- Logistic Regression

Summary of Common Mistakes in SAS

1. Handling of missing data. Check SAS output that the number of missing values is as you expect. Make sure you have changed any `NA` results in the raw data to be missing, since SAS would consider `NA` as a valid category (a non-missing character result).
2. Make sure you consider continuous or categorical variables as you intended. Just because a variable is character or numeric in the dataset, doesn't mean SAS will treat it that way in the model. You have to use Class row to tell SAS which variables should be treated as character factors. You also have to use `ref=' '` to tell SAS which is the reference category, otherwise SAS by default which use the last value of the variable alphabetically (e..g a categorical variable with 1, 2, 3 would default to 3 as the reference).
3. Be careful you are modelling the correct event (response vs non-response, or weight_gain vs weight_loss for example)
4. Be careful when interpreting any odds ratios that you have the factor of interest the correct way around (0 vs 1, or 1 vs 0)
5. If using proc logistic, be careful of how SAS creates its parameters used in the model as this determines how you can use the parameter estimates! It is often easiest to use `param=glm` so that the exp(maximum liklihood parameter estimate) = odds ratio. Check the class level information (Design variables) is as you would expect. See below for more detail on other parameterization methods.

Identifying Packages Not Being Maintained

- McNemar's Chi Squared
- {epibasix} package
  - Categorised as 'High Risk' by {riskmetrics} package
  - Unknown by the author what equation they used to calculated the CIs

Analysis	Supported in R	Supported in SAS	Results Match	Notes
McNemar's Chi-Squared test	Yes	Yes	✅	By default SAS doesn't include the continuity correction. In R use {stats} or {coin}
Cohen's Kappa CI	Yes	Yes	✅	In R use {vcd}

Finding Bugs and Getting them Corrected

- {RBesT} produces Clopper Pearson CIs
- Matched {cardx} and {Hmisc} except 2 cases

1. x = 0 (0% responders), in which case the lower limit does not match.
2. x = n (100% responders), in which case the upper limit does not match.

- Authors informed
- Package updated & re-released on CRAN as v1.8-0

- CAMIS Improves reliability & confidence in R packages

Key Impacts

- Documentation of Differences due to Methodology use
  - Continuity corrections
  - Different Defaults
  - Methods / Options not available in both software
- Common implementation mistakes and any unclear software documentation explained
- Bugs identified (and R packages updated & fixed!)
- CAMIS Is a Trusted Resource for Industry Guidance

Example

- For a study comparing treatments A and B in patients. Apply the Wilcoxon rank-sum test to analyse the change from baseline FEV1 post-mannitol challenge 2. Report the Hodges-Lehmann estimate of the median difference between treatments (B vs A), corresponding 95% CI and the associated p-value

How to Compare

Is this method possible in both software?

Are there multiple ways to do this in each software?

Are considerations we need to make when choosing a dataset?

Example

Is this method possible in both software?

Are there multiple ways to do this in each software?

- stats
- pairwiseCI
- coin

Comparing Packages

- Is this package being actively maintained?
- Is the author known in this field?
- Does one of the packages have better documentation?
- Look at community adoption?
- Can the packages handle different edge cases?

Example

pairwiseCI: Confidence Intervals for Two Sample Comparisons

Calculation of the parametric, nonparametric confidence intervals for the difference or ratio of location parameters, nonparametric confidence interval for the Behrens-Fisher problem and for the difference, ratio and odds-ratio of binomial proportions for comparison of independent samples. Common wrapper functions to split data sets and apply confidence intervals or tests to these subsets. A by-statement allows calculation of CI separately for the levels of further factors. CI are not adjusted for multiplicity.

Version:	0.1-27
Depends:	MCPAN, coin (≥ 1.3-0)
Imports:	graphics, stats, boot, MASS, mcprofile
Published:	2019-03-11
DOI:	10.32614/CRAN.package.pairwiseCI
Author:	Frank Schaarschmidt [aut, cre], Daniel Gerhard [aut]
Maintainer:	Frank Schaarschmidt <schaarschmidt at biostat.uni-hannover.de>
License:	GPL-2
NeedsCompilation:	no
CRAN checks:	pairwiseCI results

Example

coin: Conditional Inference Procedures in a Permutation Test Framework

Conditional inference procedures for the general independence problem including two-sample, K-sample (non-parametric ANOVA), correlation, censored, ordered and multivariate problems described in <doi:10.18637/jss.v028.i08>.

Version:	1.4-3
Depends:	R (≥ 3.6.0), survival
Imports:	methods, parallel, stats, stats4, utils, libcoin (≥ 1.0-9), matrixStats (≥ 0.54.0), modeltools (≥ 0.2-9), mvtnorm (≥ 1.0-5), multcomp
LinkingTo:	libcoin (≥ 1.0-9)
Suggests:	xtable, e1071, vcd, TH.data (≥ 1.0-7)
Published:	2023-09-27
DOI:	10.32614/CRAN.package.coin
Author:	Torsten Hothorn [aut, cre], Henric Winell [aut], Kurt Hornik [aut], Mark A. van de Wiel [aut], Achim Zeileis [aut]
Maintainer:	Torsten Hothorn <Torsten.Hothorn at R-project.org>
License:	GPL-2
URL:	http://coin.r-forge.r-project.org

Example

```
> cranlogs::cran_downloads(package = c("coin", "pairwiseCI"))
        date count    package
1 2025-05-25   724       coin
2 2025-05-25    11 pairwiseCI
```

Filter

	package	version	pkg_ref	pkg_score
1	coin	1.4-3	<environment: 0x17addce58>	0.9106095
2	pairwiseCI	0.1-27	<environment: 0x17b4b5240>	0.9758364

Example

Is this method possible in both software? → Are there multiple ways to do this in each software? → Are considerations we need to make when choosing a dataset?

- Because we have known issues with ties, we want to make sure to have a dataset that includes them

Example

	R {coin}	SAS
P-value (asymptotic)	0.009392	0.0094
CI (exact)	-0.76 – -0.1	-0.7600 - -0.1000
P-values (exact)	0.008181	0.0082
CI (asymptotic)	NA	-0.77 – -0.0900

Example

- For a study comparing treatments A and B in patients. Apply the Wilcoxon rank-sum test to analyse the change from baseline FEV1 post-mannitol challenge 2. Report the Hodges-Lehmann estimate of the median difference between treatments (B vs A), corresponding 95% exact CI and the associated exact p-value (without continuity correction)

CAMIS Objectives

- **Understand and Document analysis result discrepancies across software**
- **Demonstrate the methodology through examples**
- **Document in open GitHub repository**
- **Grow the repository, increasing quality and quantity of information**

Questions ?

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

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

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

phuse

The Global Healthcare Data Science Community

Contact Channels
UK +44 1843 609600
office@phuse.global
phuse.global

Social Media
@phusetwitta
/phusebook
/phusetube
/company/phuse

phuse

CAMIS