UNIVERSITY OF
LEICESTER

Analysing Results From Monte Carlo Simulation Studies

A Gasparini (@ellessenne) · TP Morris · IR White · MJ Crowther

useR! 2019

Monte Carlo simulation studies

Monte Carlo simulation studies are computer experiments that involve generating data by pseudo-random sampling.

The Process

In a simulation study, we:

The Process

In a simulation study, we:

1. Generate data from a known distribution (so that we know the “truth”);

The Process

In a simulation study, we:

1. Generate data from a known distribution (so that we know the “truth”);

2. Analyse the data;

The Process

In a simulation study, we:

1. Generate data from a known distribution (so that we know the “truth”);
2. Analyse the data;
3. Compare the analysis results with the truth.

Monte Carlo simulation studies are useful...

Monte Carlo simulation studies provide an invaluable tool for statistical and biostatistical research.

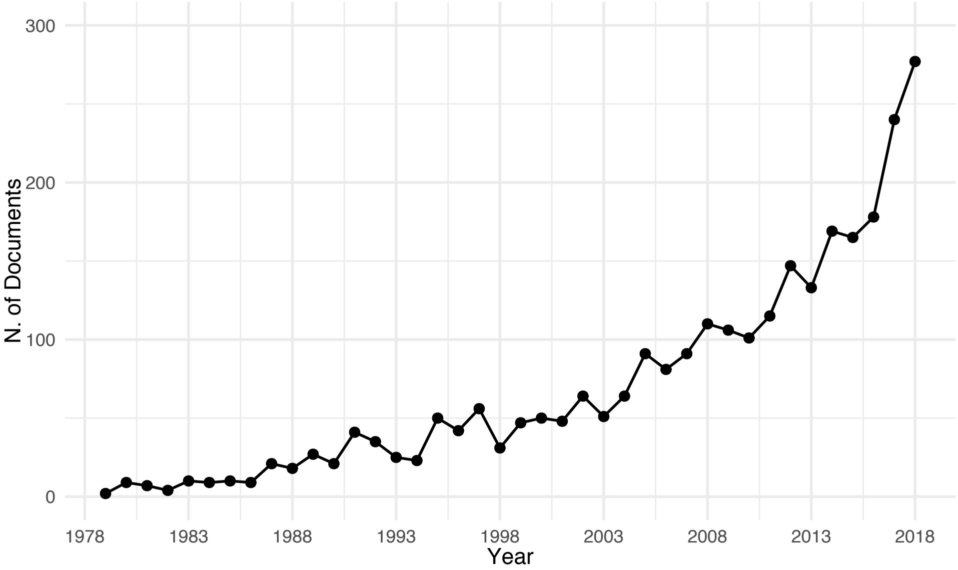
Monte Carlo simulation studies are useful...

Monte Carlo simulation studies provide an invaluable tool for statistical and biostatistical research.

They can also help to answer questions such as:

- Is an estimator biased in a finite sample?
- Do confidence intervals for a given parameter achieve the desired nominal level of coverage?
- How does a newly developed method compare to an established one?
- What is the power to detect a desired effect size under complex experimental settings and analysis methods?
- You name it!

…and increasingly popular!



Scopus search key: `TITLE-ABS-KEY ("Monte Carlo simulation study")`

There are issues...

1. Monte Carlo standard errors of performance measures are often not computed/reported;

There are issues…

1. Monte Carlo standard errors of performance measures are often not computed/reported;
2. Reproducibility of results;

There are issues...

1. Monte Carlo standard errors of performance measures are often not computed/reported;
2. Reproducibility of results;
3. Dissemination of results.

Disclaimer!

I will not cover how to plan, design, and run a Monte Carlo simulation study.

An full example on modelling survival data is included with `rsimsum`:

```
vignette(”relhaz”, package = ”rsimsum”)
```

Also, check out the tutorial paper by Morris *et al*. (2019).

Enter `rsimsum`

Why yet another R package?

Enter `rsimsum`

Why yet *another R package?*

- There is a similar package in Stata, but nothing comparable in R (that I know);
- Several performance measures are supported - no need to do tedious (and error-prone) calculations by hand;
- Monte Carlo standard errors are computed and displayed by default;
- Several quick plotting options for fast iteration/exploration.

rsimsum's main function: simsum

```
args(rsimsum::simsum)

## function (data, estvarname, true, se, methodvar = NULL, ref = NULL,
##     by = NULL, ci.limits = NULL, dropbig = FALSE, x = FALSE,
##     control = list())
## NULL
```

Documentation: https://ellessenne.github.io/rsimsum/

Example: t-test (1)

Aim: investigate the performance of the two-sample t-test when

1. Data is skewed;
2. Variances are unequal in the two groups.

Example: t-test (2)

Data-generating mechanisms (DGMs):

1. Simulating 60 observations, with 2:1 groups ratio;
2. Fully fractional design, varying equal/unequal variance, and skewed/non-skewed data.

Example: t-test (2)

Data-generating mechanisms (DGMs):

1. Simulating 60 observations, with 2:1 groups ratio;
2. Fully fractional design, varying equal/unequal variance, and skewed/non-skewed data.

Methods:

1. t-test with pooled variance;
2. t-test with non-pooled variance.

Example: t-test (2)

Data-generating mechanisms (DGMs):

1. Simulating 60 observations, with 2:1 groups ratio;
2. Fully fractional design, varying equal/unequal variance, and skewed/non-skewed data.

Methods:

1. t-test with pooled variance;
2. t-test with non-pooled variance.

Replications: 2,000 per DGM.

Case study: data

```
dplyr::glimpse(data)
```

```
## Observations: 16,000
## Variables: 8
## $ diff <dbl> -2.802464229, -0.079683569, -2.802464229, -0.079683569, -1.7...
## $ se <dbl> 1.279933, 2.218313, 1.197805, 2.947133, 1.363249, 1.723423, ...
## $ df <dbl> 58.00000, 58.00000, 46.14587, 22.18708, 58.00000, 58.00000, ...
## $ i <int> 1, 1, 1, 1, 1, 1, 1, 1, 2, 2, 2, 2, 2, 2, 2, 2, 3, 3, 3, 3, ...
## $ dgm <fct> ”Equal, Non-skewed”, ”Unequal, Non-Skewed”, ”Equal, Non-skew...
## $ method <fct> t-test (P), t-test (P), t-test (NP), t-test (NP), t-test
(P)...
## $ dist <fct> N, N, N, N, Gamma, Gamma, Gamma, Gamma, N, N, N, N, Gamma, G...
## $ var <fct> Equal, Unequal, Equal, Unequal, Equal, Unequal, Equal, Unequ...
```

Summarising a simulation study (1)

```
s <- rsimsum::simsum(
  data = data, estvarname = ”diff”, se = ”se”, true = -1, methodvar = ”method”,
  by = ”dgm”, ref = ”t-test (NP)”, x = TRUE
)
s
```

```
## Summary of a simulation study with a single estimand.
##
## Method variable: method
##  Unique methods: t-test (NP), t-test (P)
##  Reference method: t-test (NP)
##
## By factors: dgm
##
## Monte Carlo standard errors were computed.
```

Summarising a simulation study (2)

```
summary(s, stats = ”bias”)

## Values are:
##  Point Estimate (Monte Carlo Standard Error)
##
## Bias in point estimate:
##                   dgm      t-test (NP)       t-test (P)
##    Equal, Non-skewed  0.0054 (0.0302)  0.0054 (0.0302)
##  Unequal, Non-Skewed -0.0461 (0.0547) -0.0461 (0.0547)
##        Equal, Skewed -0.0610 (0.0300) -0.0610 (0.0300)
##      Unequal, Skewed -0.0258 (0.0533) -0.0258 (0.0533)
```

Summarising a simulation study (3)

```
summary(s, stats = ”cover”)
```

```
## Values are:
##  Point Estimate (Monte Carlo Standard Error)
##
## Coverage of nominal 95% confidence interval:
##                   dgm     t-test (NP)      t-test (P)
##     Equal, Non-skewed 0.9590 (0.0044) 0.9600 (0.0044)
##   Unequal, Non-Skewed 0.9350 (0.0055) 0.8725 (0.0075)
##         Equal, Skewed 0.9495 (0.0049) 0.9525 (0.0048)
##       Unequal, Skewed 0.9290 (0.0057) 0.8770 (0.0073)
```

Performance measures implemented in `rsimsum`

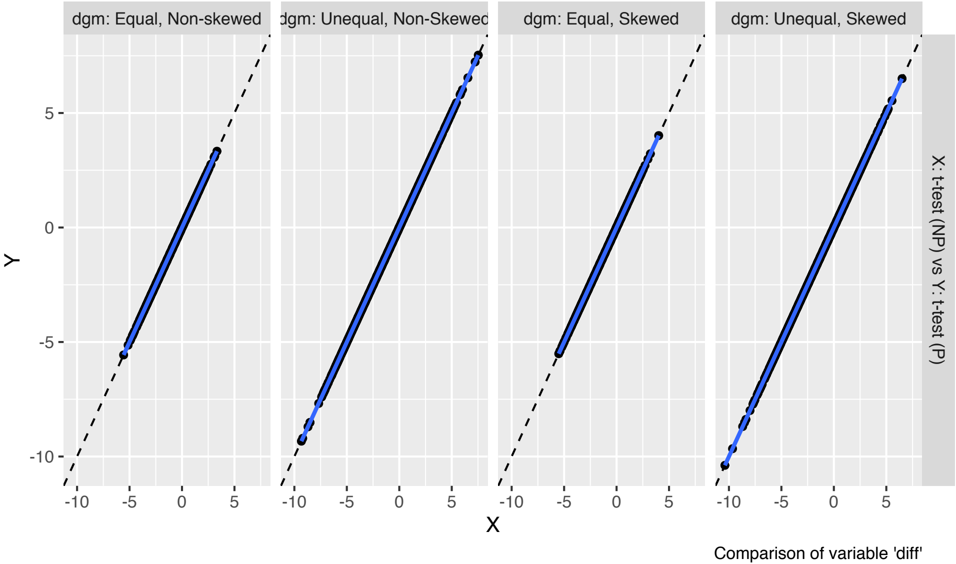
The following performance measures are implemented in `rsimsum`:

- Bias;
- Empirical SE, relative % increase in precision, model-based SE, and relative % error in model-based SE;
- Mean squared error (MSE);
- Coverage probability and bias-corrected coverage probability;
- Power of type I error.

Each performance measure is described in more detail elsewhere (Morris *et al.*, 2019).

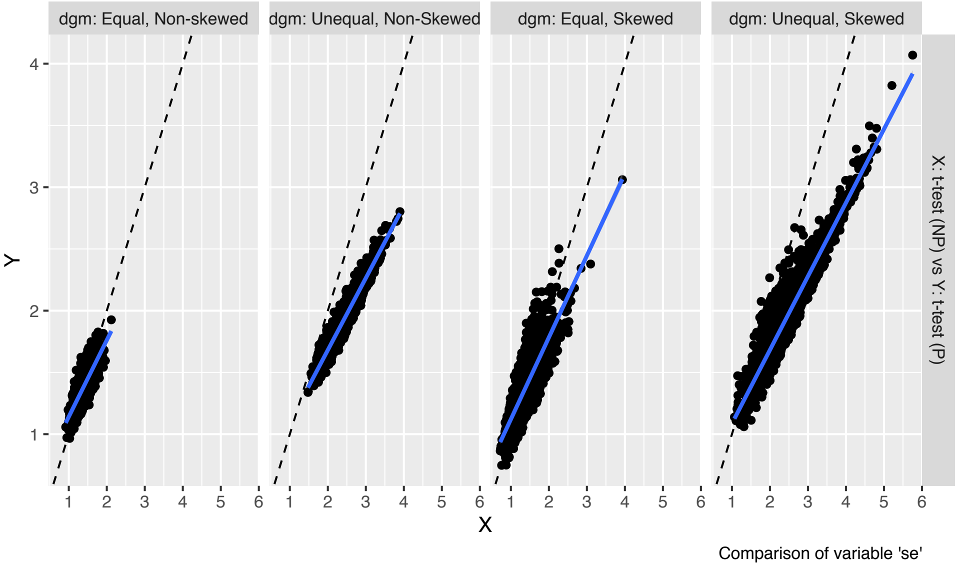
Plotting point estimates

```
autoplot(object = s, type = "est")
```



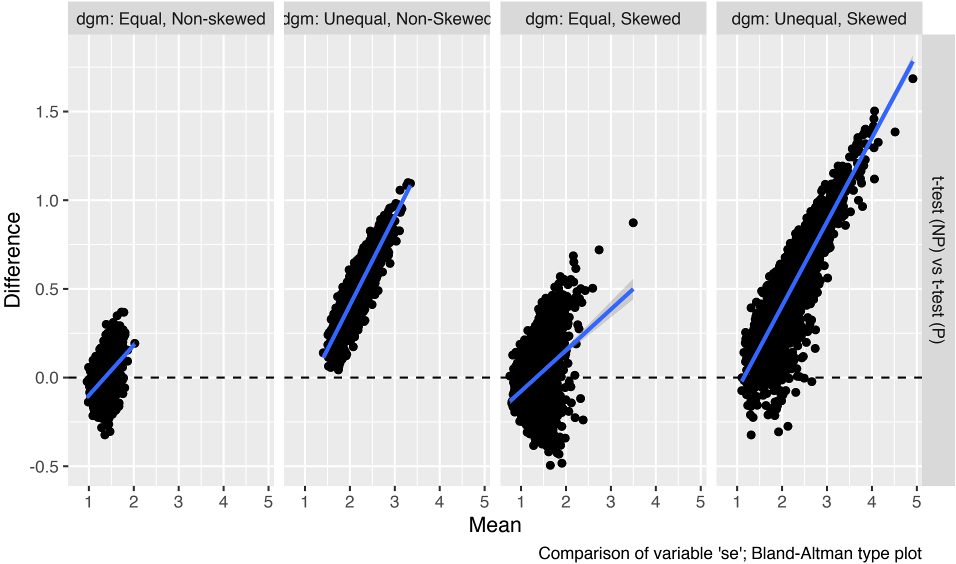
Plotting standard errors

```
autoplot(object = s, type = "se")
```



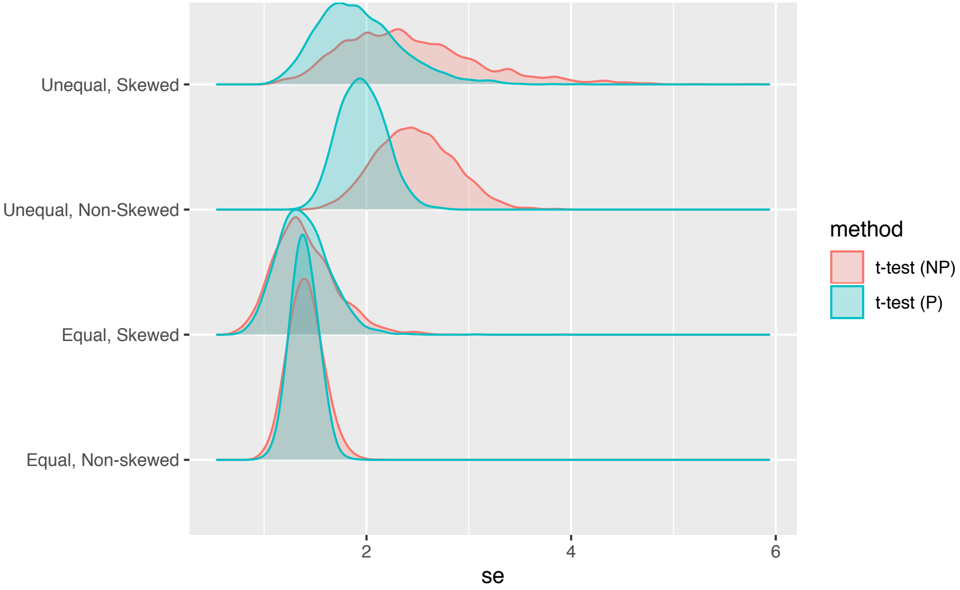
Plotting standard errors

```
autoplot(object = s, type = ”se_ba”)
```



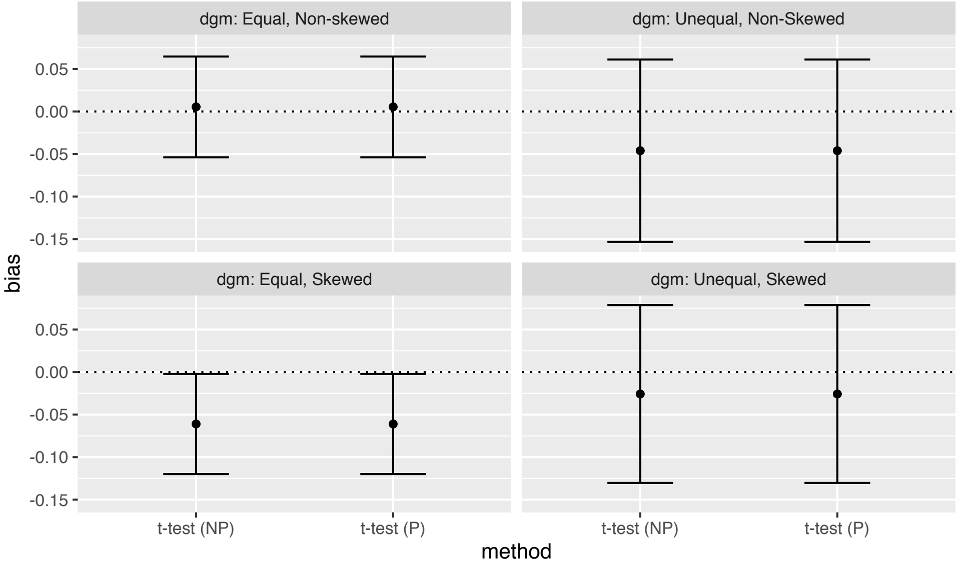
Plotting standard errors

```
autoplot(object = s, type = "se_ridge")
```



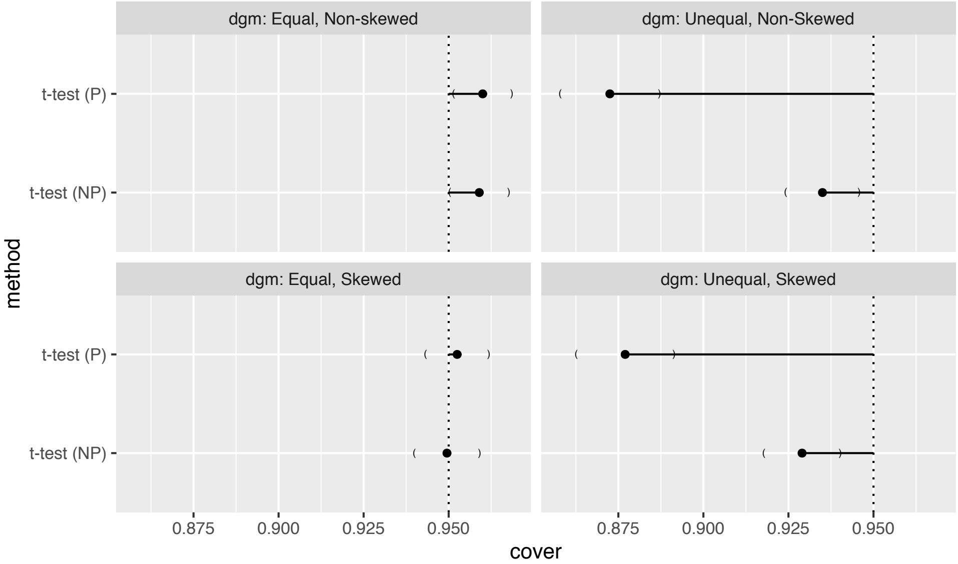
Plotting results

```
autoplot(object = summary(s), type = "forest", stats = "bias")
```



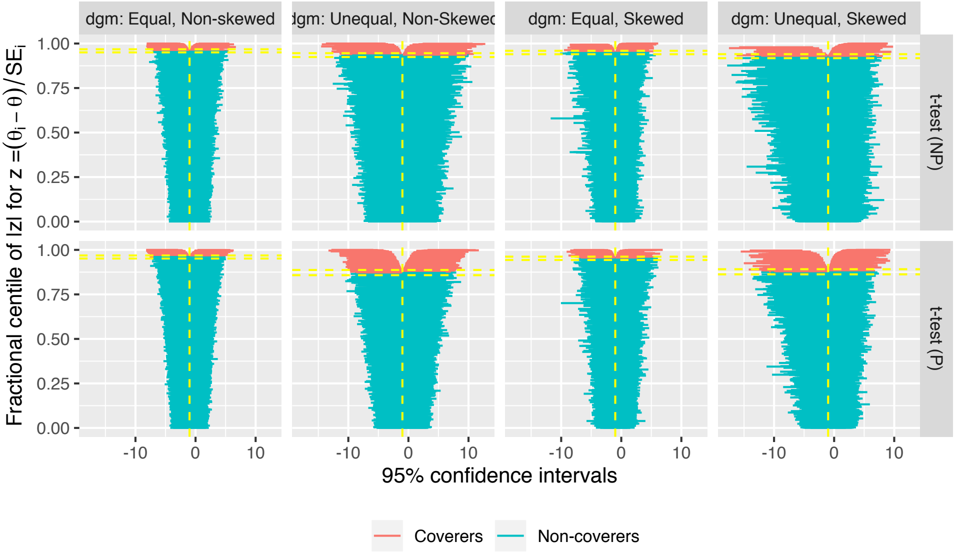
Plotting results

```
autoplot(object = summary(s), type = "lolly", stats = "cover")
```



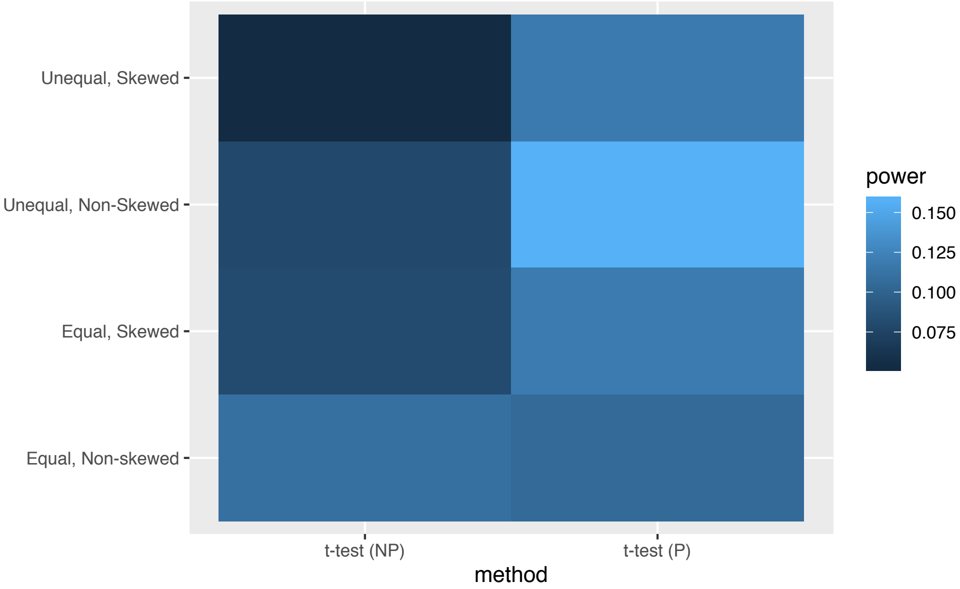
Plotting results

```
autoplot(object = s, type = "zip")
```



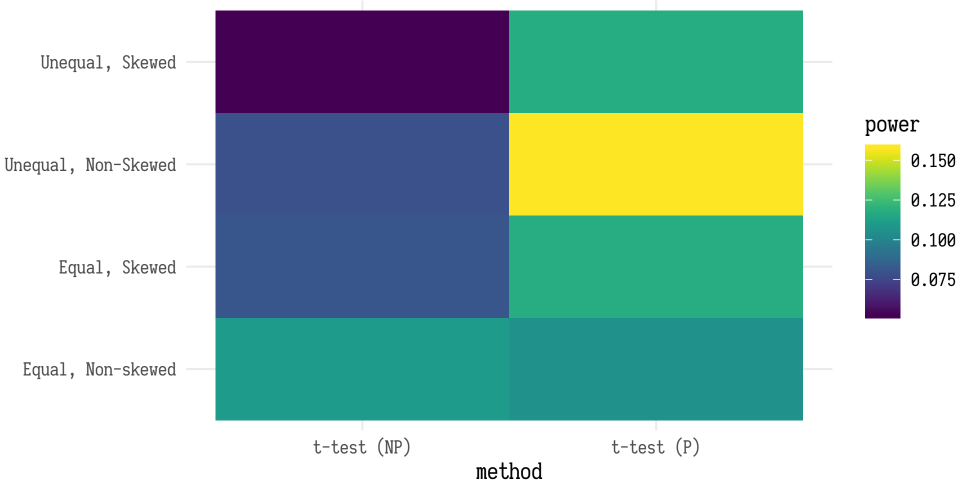
Plotting results

```
autoplot(object = s, type = "heat", stats = "power")
```



Plotting results

```
autoplot(object = s, type = "heat", stats = "power") +
  viridis::scale_fill_viridis() +
  ggplot2::theme_minimal(base_family = "Iosevka Slab")
```



Multiple estimands

```
args(rsimsum::multisimsum)
```

```
## function (data, par, estvarname, true, se, methodvar = NULL,
##     ref = NULL, by = NULL, ci.limits = NULL, dropbig = FALSE,
##     x = FALSE, control = list())
## NULL
```

INTEREST

INteractive **T**ool for **E**xploring **RE**sults from **S**imulation s**T**udies

INTEREST: Why?

Why a Shiny app?

INTEREST: Why?

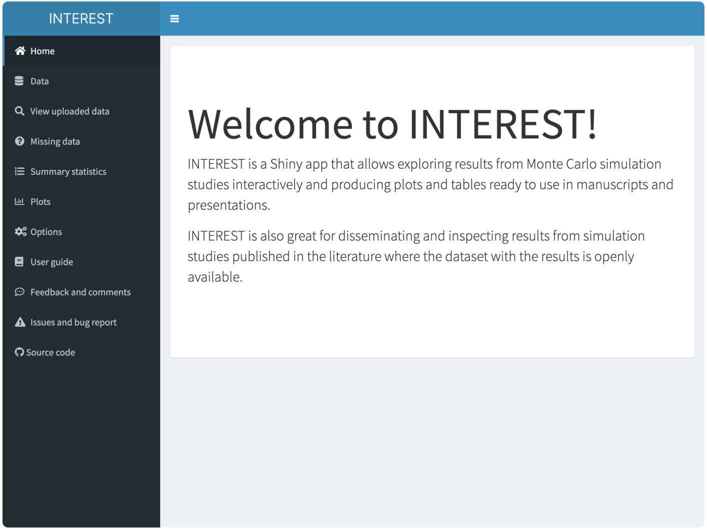
Why a Shiny app?

- Dissemination of results and open science;
- Fast(er) iteration and exploration of results;
- Supporting devices where R does not run natively (smartphones, Chromebooks, ...).

Dissemination is key

1. Can drive practitioners and applied statisticians to methods that have been shown to perform well in their practical settings;
2. Can guide researchers to develop new methods in promising directions;
3. Can provide insights into less established methods.

INTEREST: Homepage

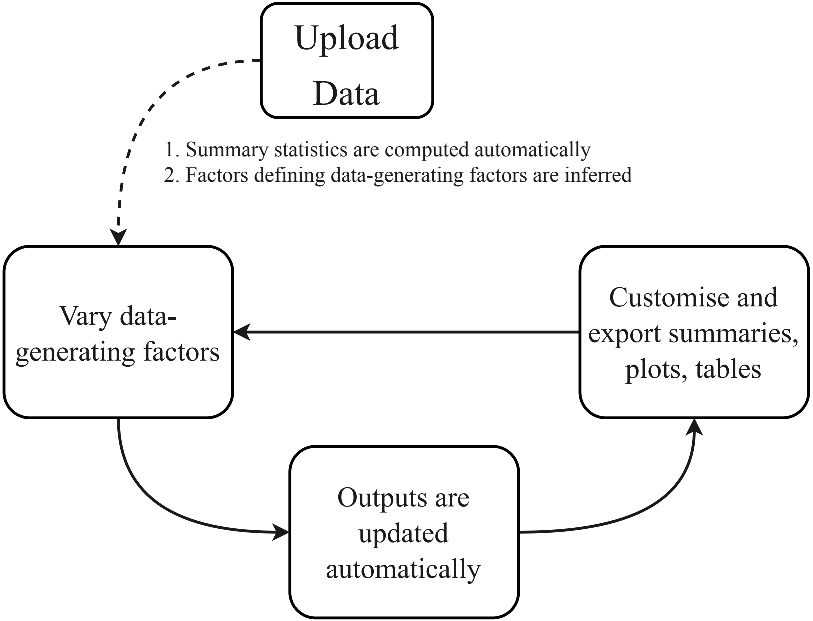


Welcome to INTEREST!

INTEREST is a Shiny app that allows exploring results from Monte Carlo simulation studies interactively and producing plots and tables ready to use in manuscripts and presentations.

INTEREST is also great for disseminating and inspecting results from simulation studies published in the literature where the dataset with the results is openly available.

INTEREST: Workflow


Upload Data
1. Summary statistics are computed automatically
2. Factors defining data-generating factors are inferred
Vary data-generating factors
Customise and export summaries, plots, tables
Outputs are updated automatically

INTEREST: Demo

`http://interest.shinyapps.io/interest/`

How to get rsimsum and INTEREST?

rsimsum can be installed directly from CRAN:

```
install.packages(”rsimsum”)
# Development version on GitHub:
# require(”remotes”)
# remotes::install_github(repo = ”ellessenne/rsimsum”)
```

INTEREST is on GitHub:

```
# require(”remotes”)
remotes::install_github(repo = ”ellessenne/interest”)
```

What is coming next?

`rsimsum`:

- Nested loop plot for simulation studies with several DGMs (Rücker and Schwarzer, 2014);
- Methods to easily reproduce plots generated by `autoplot`;
- Methods to directly export (pretty) LaTeX tables;
- ...

INTEREST:

- Additional exporting tools for INTEREST;
- Support for simulation studies with multiple estimands in INTEREST;
- ...

Thank you!

References:

- *Using simulation studies to evaluate statistical methods*. Morris TP, White IR, and Crowther MJ (2019). Statistics in Medicine 38(11):2074–2102, DOI: 10.1002/sim.8086
- *rsimsum: Summarise results from Monte Carlo simulation studies*. Gasparini A (2018). Journal of Open Source Software 3(26):739, DOI: 10.21105/joss.00739
- `rsimsum`'s website: `https://ellessenne.github.io/rsimsum/`

Slides available online: `https://tinyurl.com/useR-2019`