

Everything on the Table

Demystifying the Results of Simulation Studies and Other Complex Tables

Matthew Sigal
York University

February 29th, 2016



Introduction

This presentation will outline my current dissertation research.

Primary Goals

- ▶ Introduce Monte Carlo designs
 - ▶ What? Why? How?
 - ▶ How are results typically presented?
- ▶ Use this as a springboard for a discussion on how such presentations might be improved.
- ▶ Provide an R package, `SimDisplay` that integrates with `SimDesign` and extends `ggplot2` to aid in this endeavor.

This project is currently in the proposal stage, so feedback would be greatly appreciated!

Monte Carlo Designs. . .

Monte Carlo Simulation Studies (MCSS) provide a pivotal foundation for research in Quantitative Psychology, and in applied statistics at large.

They have been used to:

- ▶ Assess the performance of the RMSEA in SEM models with small degrees of freedom (Kenny et al., 2015)

Monte Carlo Designs. . .

Monte Carlo Simulation Studies (MCSS) provide a pivotal foundation for research in Quantitative Psychology, and in applied statistics at large.

They have been used to:

- ▶ Assess the performance of the RMSEA in SEM models with small degrees of freedom (Kenny et al., 2015)
- ▶ Examine the effects of skewness and kurtosis in linear mixed models (Arnau et al., 2013)

Monte Carlo Designs. . .

Monte Carlo Simulation Studies (MCSS) provide a pivotal foundation for research in Quantitative Psychology, and in applied statistics at large.

They have been used to:

- ▶ Assess the performance of the RMSEA in SEM models with small degrees of freedom (Kenny et al., 2015)
- ▶ Examine the effects of skewness and kurtosis in linear mixed models (Arnau et al., 2013)
- ▶ Investigate the behaviour of statistics (such as Pearson's R) at various sample sizes (Schönbrodt and Perugini, 2013)

... are here to stay!

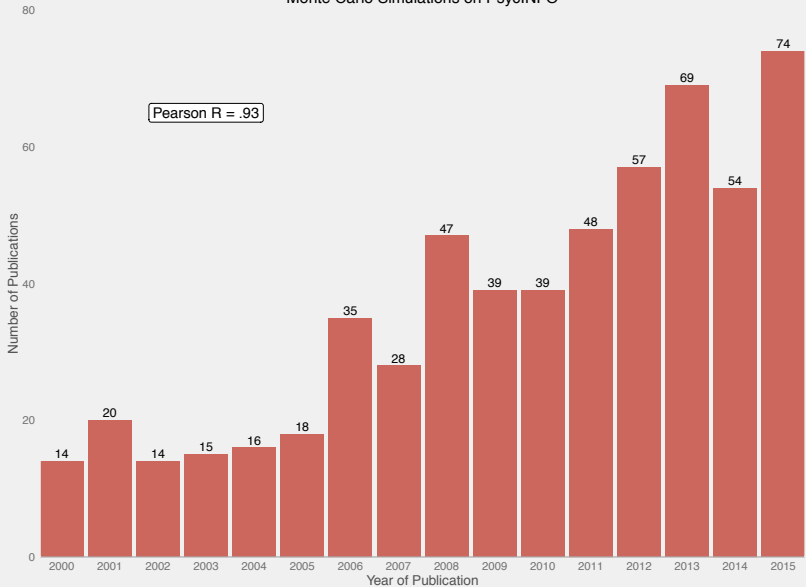
MCSS are used in a wide variety of fields and for many purposes:

- ▶ To help estimate if lower income areas have more pedestrian casualties (Noland et al., 2013)
- ▶ To create projections of teen pregnancy rates (Sayegh et al., 2010)

Within Psychology

A cursory search for **peer reviewed** articles using the search string `all("Monte Carlo Simulation")` in **scholarly journals** on **PsycINFO** alone yielded an evident relationship...

Monte Carlo Simulations on PsycINFO



Within Psychology

MCSS are especially prevalent in the pages of *Multivariate Behavioral Research* and *Structural Equation Modeling (SEM)*. In fact, these two journals have printed essential guides for researchers:

- ▶ Skrondal (2000): Design and analysis of Monte Carlo experiments: Attacking the conventional wisdom

Within Psychology

MCSS are especially prevalent in the pages of *Multivariate Behavioral Research* and *Structural Equation Modeling (SEM)*. In fact, these two journals have printed essential guides for researchers:

- ▶ Skrondal (2000): Design and analysis of Monte Carlo experiments: Attacking the conventional wisdom
- ▶ Paxton, Curran, Bollen, et al. (2001): Monte Carlo experiments: Design and implementation

Within Psychology

MCSS are especially prevalent in the pages of *Multivariate Behavioral Research* and *Structural Equation Modeling (SEM)*. In fact, these two journals have printed essential guides for researchers:

- ▶ Skrondal (2000): Design and analysis of Monte Carlo experiments: Attacking the conventional wisdom
- ▶ Paxton, Curran, Bollen, et al. (2001): Monte Carlo experiments: Design and implementation
- ▶ Boomsma (2013): Reporting Monte Carlo studies in *Structural Equation Modeling*

Monte Carlo Simulation Studies

What are Monte Carlo Simulation Studies? Why do we care?

MCSS are **experiments** in which certain parameters, which are known and fixed by the researcher, are used to generate random data and then estimate the behavior of other statistics. This is repeated over many *iterations*, and across many *conditions*.

Monte Carlo Simulation Studies

For example, we could:

- ▶ Evaluate the performance of a new statistic
- ▶ Investigate the performance of a statistic under various assumption violations
- ▶ Determine Power or Type I error rates for various designs
- ▶ Determine how well we recover parameters under various conditions
- ▶ Simulate 'realistic' data to address hard to study phenomena

Monte Carlo Simulation Studies

For SEM in particular, MCSS are an excellent approach for evaluating estimators and goodness-of-fit statistics under a variety of conditions, model complexity, and model misspecification.

As stated by Paxton et al., "... many topics in SEM would benefit from an empirical analysis through Monte Carlo methods" (2001, p. 288).

Conducting MCSS Research

1. Develop a theoretically derived research question
2. Create a valid model
3. Design specific experimental conditions
4. Choose values for the population parameters
5. Choose an appropriate software package
6. Execute the simulation and repeat
7. Determine a file storage solution
8. Troubleshoot and verify
9. Summarize results

Brief Example

Can we use likelihood difference approach to detect differential item functioning for multi-dimensional IRT models?¹

- ▶ Investigate **false detection rates** (FDR, or type I DIF tests for detecting no difference between groups), where the underlying item structure for all items is equivalent across the two groups.
- ▶ Investigate **power**, where a true difference exists for the slope of one or more particular items across the groups on one of the underlying factors.

¹Sigal and Chalmers, forthcoming.

Brief Example

Four design variables were explored to determine their effect on detection rates:

- ▶ number of items (10, 15, 30)
- ▶ sample size per group (250, 500, 1000, 2000)
- ▶ number of items with DIF (0, 1, 2, or 3)
- ▶ for the conditions with DIF, also manipulated the amount of DIF (medium or small)

Final Design

After removing redundant cells, we have 12 FDR and 72 DIF conditions: 84 total cells.

MCSS Presentation

How are results from such designs presented? Are they done in a meaningful way?

In previous work, methods for conveying the findings from MCSS have been given little attention.

For instance, Paxton et al. (2001) state that results can be presented descriptively, graphically, and inferentially but provide little detail on how to do so.

MCSS Presentation

“...reading results from Monte Carlo studies in whatever form should be a revelatory task, not a baffling puzzlement.”

–Boomsma, 2013, p. 534.

Prototypical Results

The following table is from Ramsey & Ramsey (2009), presented in the *British Journal of Mathematical and Statistical Psychology*.

Straight forward simulation design:

- ▶ Goal:
 - ▶ to compare the performance of 10 pairwise multiple comparison procedures (MCPs)
- ▶ Design:
 - ▶ degree of heteroskedasticity (c , equal variance, and multiplied by 2, 4, and 10)
 - ▶ number of groups (k , from 4 to 8)
 - ▶ sample size per group (n , from 2 to 500)
- ▶ Primary output:
 - ▶ Type I error rates from full true null models

Table 1. Type I errors for 10 MCPs at $\alpha = .05$, $k = 4$, equal n_i , variance multiplier c and a true full null hypothesis

n	T3	C	GH	GF*	PF*	GH9	GH8	GH7	GH6	GH5
(a) $c = 1$										
2	.0791	.0117	.0445	.0048	.0039	.0368	.0313	.0260	.0197	.0158
3	.0397	.0135	.0512	.0137	.0171	.0439	.0385	.0331	.0272	.0216
4	.0391	.0157	.0514	.0186	.0256	.0450	.0388	.0335	.0283	.0234
5	.0383	.0173	.0499	.0216	.0311	.0446	.0379	.0328	.0265	.0218
6	.0426	.0231	.0536	.0286	.0371	.0472	.0426	.0375	.0309	.0253
7	.0388	.0239	.0497	.0259	.0351	.0434	.0386	.0338	.0295	.0252
8	.0446	.0301	.0534	.0305	.0373	.0479	.0443	.0391	.0342	.0292
9	.0410	.0263	.0511	.0301	.0390	.0460	.0407	.0360	.0290	.0238
10	.0471	.0343	.0562	.0363	.0441	.0518	.0468	.0410	.0353	.0291
11	.0417	.0313	.0531	.0326	.0423	.0464	.0412	.0368	.0311	.0256
12	.0423	.0333	.0518	.0342	.0433	.0470	.0422	.0376	.0317	.0261
13	.0446	.0371	.0545	.0367	.0440	.0488	.0439	.0390	.0344	.0295
14	.0439	.0367	.0525	.0338	.0436	.0474	.0437	.0381	.0333	.0274
21	.0430	.0401	.0517	.0387	.0479	.0473	.0417	.0369	.0321	.0269
24	.0422	.0406	.0510	.0383	.0475	.0469	.0414	.0367	.0323	.0284
28	.0405	.0400	.0496	.0371	.0467	.0448	.0397	.0354	.0301	.0249
29	.0402	.0391	.0472	.0356	.0426	.0436	.0387	.0333	.0288	.0236
30	.0416	.0408	.0510	.0398	.0498	.0456	.0402	.0361	.0305	.0248
31	.0436	.0432	.0525	.0406	.0478	.0477	.0418	.0369	.0311	.0257
32	.0427	.0425	.0513	.0388	.0467	.0474	.0411	.0365	.0320	.0269
33	.0365	.0364	.0457	.0333	.0425	.0394	.0352	.0312	.0260	.0214
34	.0467	.0469	.0546	.0430	.0517	.0506	.0460	.0413	.0360	.0301
35	.0409	.0413	.0488	.0361	.0434	.0436	.0397	.0344	.0287	.0234
36	.0431	.0440	.0527	.0402	.0480	.0463	.0418	.0367	.0322	.0264
37	.0412	.0420	.0495	.0377	.0442	.0448	.0395	.0359	.0310	.0249
38	.0423	.0439	.0521	.0375	.0459	.0465	.0408	.0351	.0295	.0246
39	.0446	.0461	.0545	.0403	.0485	.0485	.0435	.0381	.0325	.0274
40	.0405	.0425	.0487	.0383	.0467	.0445	.0394	.0353	.0290	.0245
50	.0496	.0520	.0581	.0457	.0528	.0534	.0481	.0416	.0356	.0303
60	.0417	.0448	.0506	.0391	.0460	.0452	.0402	.0358	.0314	.0259
70	.0401	.0438	.0482	.0384	.0462	.0437	.0392	.0340	.0284	.0236
80	.0437	.0491	.0534	.0440	.0517	.0486	.0430	.0376	.0328	.0277
100	.0431	.0484	.0517	.0415	.0484	.0469	.0419	.0371	.0324	.0269
200	.0421	.0498	.0515	.0419	.0500	.0462	.0411	.0363	.0318	.0283
300	.0430	.0489	.0499	.0423	.0500	.0464	.0421	.0374	.0317	.0265
400	.0392	.0463	.0464	.0395	.0464	.0426	.0382	.0346	.0300	.0251
500	.0450	.0534	.0537	.0452	.0517	.0480	.0439	.0394	.0344	.0296
Max	.0791	.0534	.0581	.0457	.0528	.0534	.0481	.0416	.0360	.0303
(b) $c = 2$										
2	.0898	.0109	.0512	.0073	.0056	.0407	.0321	.0253	.0209	.0167
3	.0451	.0158	.0585	.0156	.0199	.0515	.0430	.0367	.0310	.0259
4	.0478	.0225	.0598	.0240	.0324	.0531	.0469	.0402	.0357	.0292
5	.0467	.0247	.0586	.0267	.0418	.0519	.0460	.0405	.0345	.0282
6	.0448	.0275	.0563	.0283	.0440	.0502	.0444	.0384	.0324	.0270
7	.0433	.0288	.0543	.0293	.0438	.0491	.0431	.0382	.0329	.0271
14	.0383	.0348	.0449	.0261	.0443	.0420	.0379	.0335	.0280	.0226
21	.0427	.0424	.0525	.0298	.0485	.0475	.0418	.0370	.0312	.0263

Table 1. (Continued)

n	T3	C	GH	GF*	PF*	GH9	GH8	GH7	GH6	GH5
25	.0391	.0394	.0456	.0267	.0430	.0411	.0380	.0337	.0295	.0250
26	.0418	.0437	.0506	.0299	.0465	.0457	.0413	.0365	.0315	.0249
27	.0424	.0435	.0527	.0320	.0472	.0470	.0413	.0354	.0317	.0262
28	.0393	.0408	.0465	.0269	.0439	.0434	.0382	.0336	.0289	.0235
29	.0398	.0409	.0481	.0285	.0447	.0434	.0388	.0348	.0296	.0246
30	.0395	.0410	.0470	.0269	.0426	.0419	.0383	.0333	.0287	.0238
31	.0427	.0434	.0497	.0299	.0466	.0451	.0417	.0364	.0326	.0279
32	.0393	.0418	.0490	.0280	.0457	.0436	.0387	.0339	.0297	.0247
33	.0402	.0422	.0479	.0286	.0462	.0440	.0393	.0353	.0298	.0249
34	.0396	.0414	.0471	.0256	.0429	.0430	.0378	.0341	.0290	.0249
35	.0425	.0450	.0498	.0303	.0490	.0457	.0411	.0362	.0313	.0266
Max	.0898	.0450	.0598	.0320	.0490	.0531	.0469	.0405	.0357	.0292
(c) c = 4										
2	.1195	.0179	.0689	.0133	.0120	.0566	.0464	.0387	.0322	.0267
3	.0618	.0291	.0742	.0241	.0306	.0670	.0600	.0520	.0457	.0397
4	.0513	.0299	.0623	.0238	.0334	.0567	.0504	.0440	.0385	.0328
5	.0502	.0341	.0609	.0246	.0378	.0550	.0494	.0437	.0382	.0311
6	.0436	.0317	.0547	.0203	.0364	.0489	.0432	.0355	.0308	.0270
7	.0397	.0330	.0489	.0219	.0390	.0440	.0396	.0350	.0302	.0254
8	.0402	.0336	.0507	.0213	.0379	.0450	.0401	.0350	.0296	.0247
9	.0419	.0378	.0516	.0242	.0410	.0469	.0416	.0365	.0309	.0261
10	.0390	.0371	.0494	.0233	.0427	.0440	.0389	.0341	.0289	.0248
11	.0404	.0379	.0479	.0205	.0369	.0443	.0400	.0350	.0297	.0253
12	.0388	.0377	.0467	.0210	.0403	.0419	.0384	.0342	.0291	.0249
13	.0388	.0387	.0480	.0220	.0423	.0436	.0386	.0349	.0298	.0236
14	.0395	.0395	.0482	.0206	.0400	.0434	.0391	.0350	.0303	.0247
15	.0365	.0370	.0454	.0208	.0413	.0403	.0361	.0307	.0267	.0218
16	.0335	.0351	.0414	.0186	.0354	.0378	.0328	.0290	.0253	.0212
17	.0359	.0386	.0443	.0189	.0361	.0405	.0356	.0311	.0269	.0229
18	.0380	.0398	.0460	.0201	.0409	.0414	.0378	.0322	.0279	.0236
21	.0392	.0419	.0473	.0203	.0417	.0423	.0388	.0341	.0280	.0229
28	.0337	.0373	.0424	.0187	.0381	.0378	.0325	.0273	.0228	.0188
35	.0366	.0396	.0422	.0196	.0398	.0391	.0357	.0316	.0273	.0228
Max	.1195	.0450	.0742	.0320	.0490	.0670	.0600	.0520	.0457	.0397
(d) c = 10										
2	.1555	.0351	.0964	.0268	.0266	.0823	.0708	.0595	.0520	.0431
3	.0686	.0434	.0831	.0285	.0359	.0747	.0660	.0602	.0523	.0440
4	.0530	.0389	.0641	.0250	.0381	.0581	.0516	.0460	.0400	.0328
5	.0476	.0406	.0578	.0224	.0346	.0527	.0466	.0419	.0367	.0306
6	.0417	.0381	.0525	.0215	.0343	.0473	.0414	.0362	.0309	.0260
7	.0394	.0385	.0492	.0181	.0319	.0435	.0390	.0343	.0293	.0243
8	.0371	.0370	.0469	.0184	.0340	.0416	.0371	.0323	.0280	.0241
9	.0327	.0351	.0439	.0180	.0359	.0382	.0325	.0287	.0250	.0207
10	.0358	.0379	.0441	.0173	.0346	.0401	.0357	.0314	.0266	.0226
11	.0338	.0371	.0431	.0180	.0364	.0375	.0337	.0304	.0266	.0226
12	.0350	.0379	.0430	.0179	.0358	.0389	.0350	.0310	.0263	.0226
13	.0327	.0368	.0414	.0172	.0363	.0376	.0327	.0288	.0247	.0206
14	.0357	.0390	.0433	.0193	.0349	.0390	.0356	.0300	.0256	.0218
21	.0319	.0369	.0398	.0152	.0354	.0363	.0312	.0283	.0242	.0205

... and it is still going!

Table 1. (Continued)

<i>n</i>	T3	C	GH	GF*	PF*	GH9	GH8	GH7	GH6	GH5
27	.0315	.0369	.0397	.0161	.0352	.0347	.0312	.0277	.0230	.0196
Max	.1555	.0434	.0964	.0285	.0381	.0823	.0708	.0602	.0523	.0440
Max	.1555	.0534	.0964	.0457	.0528	.0823	.0708	.0602	.0523	.0440

Note. T3, Dunnett; C, Cochran; GH, Games–Howell; GF*, Games–Howell and Brown–Forsythe F; PF*, Peritz–Brown–Forsythe F; GH9, GH applied at 0.9α ; GH8, GH at 0.8α ; GH7, GH at 0.7α ; GH6, GH at 0.6α ; GH5, GH at 0.5α .

Even so, this table ignores:

- ▶ Many of the sample size comparisons
 - ▶ none of the (many) sample size conditions that pertain to unequal groups
- ▶ The number of groups factor. . .
 - ▶ this entire table only refers to $k = 4!$

So...

- ▶ MCSS designs are important.
- ▶ Presenting results via tables is the classic approach
- ▶ ... however, presenting the results in this manner might not be very illuminating.

Dissertation

- ▶ Primary goal is to catalogue and improve methods for finding trends in complex tables, such as those from MCSS designs
- ▶ Take principals from information design theory
- ▶ Apply these findings across three broad categories of approaches:
 - ▶ Improving Static Tables
 - ▶ Graphical Approaches
 - ▶ Interactive Approaches
- ▶ Provide an R package, `SimDisplay` that integrates with `SimDesign` to make such presentations easy.

Static Methods

Static tables can convey substantial amounts of information and should still be recognized as a suitable method for presenting MCSS results, if thoughtfully constructed.

Recommendations

- ▶ Use emphasis to signify important results (bold, italics)
- ▶ If possible, use color for added impact
- ▶ Present the rows and columns in a meaningful way
 - ▶ Use effects ordering and/or rotation (e.g., manipulate rows using clustering algorithms)

Table 1: ... and this is only 12 of the 72 rows!

N	Items	nDIF	sDIF	Average P	Percent < .05
250	10	2	1	0.13	0.66
500	10	2	1	0.08	0.8
1000	10	2	1	0.05	0.87
2000	10	2	1	0.03	0.93
250	15	2	1	0.09	0.76
500	15	2	1	0.05	0.85
1000	15	2	1	0.04	0.9
2000	15	2	1	0.02	0.95
250	30	2	1	0.07	0.82
500	30	2	1	0.05	0.87
1000	30	2	1	0.03	0.93
2000	30	2	1	0.02	0.96

Detection rates by sample size and number of items in the situation where 2 items should exhibit large DIF.

Table 2: Sorted by Sample Size

N	Items	nDIF	sDIF	Average P	Percent < .05
250	10	2	1	0.13	0.66
250	15	2	1	0.09	0.76
250	30	2	1	0.07	0.82
500	10	2	1	0.08	0.8
500	15	2	1	0.05	0.85
500	30	2	1	0.05	0.87
1000	10	2	1	0.05	0.87
1000	15	2	1	0.04	0.9
1000	30	2	1	0.03	0.93
2000	10	2	1	0.03	0.93
2000	15	2	1	0.02	0.95
2000	30	2	1	0.02	0.96

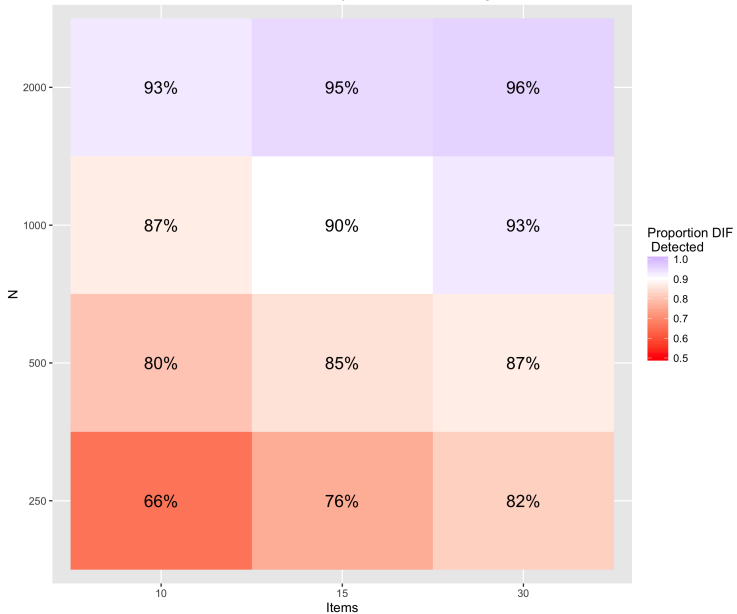
Graphical Methods

As with many statistical analyses, graphical methods should be utilized for exploring MCSS results.

Such methods range from stylized presentation of the raw table:

- ▶ e.g., heatmaps/tableplots with exact values overlaid
- ▶ Created with sensible values for shading
- ▶ For designs with 3 or more dimensions, utilize marginal tables or facet to see various relationships

DIF Detection for N by Items, with Shading



Graphical Methods

Considering how large these tables can be, it is useful to think of graphical approaches that make it easier to see what is happening.

- ▶ Extend the `tableplot` framework (Kwan et al., 2009)
 - ▶ semi-graphic display that uses both numbers and symbols (with sizes, colors, shapes providing additional information)
 - ▶ originally designed for EFA/CFA, but underlying idea can be extended.
- ▶ Treat simulation conditions as ANOVA designs
 - ▶ so-called 'meta-models'
 - ▶ utilize effect plots for the design variables
- ▶ Utilize dimension reduction techniques, such as hypothesis-error (HE) plots (Friendly, 2007)
 - ▶ Design cells presented as points on graphic

Interactive Methods

The above methods are worth pursuing for showing static slices of our results, but it would be better to be able to manipulate the data on the fly.

Interactive displays allow for the dynamic exploration of complex tables and figures. Such a display could allow for changes in:

- ▶ sorting and filtering of variables
- ▶ incorporating shading or altering the scale
- ▶ live search
- ▶ pop-up text (e.g., cell identification on HE plot)

SimDisplay

Some of these ideas have already been implemented in `SimDisplay`.

Utility Functions:

- ▶ `get_design_levels()` and `get_sim_levels()`: Extract design and result vectors from dataframe.
- ▶ `sim_design_check()`: Check properties of `SimDesign` dataframe and return for analysis

Core Features:

- ▶ `convert_df()`: Convert a dataframe to a `SimDesign` object.
- ▶ `generic_plot()` method: quick access to various faceted views of a dataframe
- ▶ `fable()` and `xtable()` methods: formatted output for static numerical tables (either to console or to LaTeX)
- ▶ `tableplot()` method to generate static semi-graphic displays
- ▶ `shinySim()` takes a `SimDesign` dataframe and creates an interactive applet.

SimDisplay

SimDisplay includes a dataframe object, `tsimresults`, with the results of a small simulation study on how the t -statistic behaves under various sample size, group size, and standard deviation ratio conditions, when evaluated with and without the Welch correction.

```
devtools::install_github("mattsigal/SimDisplay")
library(SimDisplay)
data(tsimresults)
str(tsimresults)
head(tsimresults)
library(shiny, quietly = TRUE)
runApp(shinySim(tsimresults))
```

Thank you!

Slides available at: www.matthewsigal.com/teaching

- Arnau, J., Bendayan, R., Blanca, M. J., and Bono, R. (2013). The effects of skewness and kurtosis on the robustness of linear mixed models. *Behavioural Research*, 45(3):873–879.
- Boomsma, A. (2013). Reporting Monte Carlo studies in Structural Equation Modeling. *Structural Equation Modeling*, 20:518–540.
- Friendly, M. (2007). HE plots for multivariate linear models. *Journal of Computational and Graphical Statistics*, 16(2):421–444.
- Kenny, D. A., Kaniskan, B., and McCoach, D. B. (2015). The performance of RMSEA in models with small degrees of freedom. *Sociological Methods & Research*, 44(3):486–507.
- Kwan, E., Lu, I. R. R., and Friendly, M. (2009). Tableplot: A new tool for assessing precise predictions. *Zeitschrift für Psychologie / Journal of Psychology*, 217(1):38–48.
- Noland, R. B., Klein, N. J., and Tulach, N. K. (2013). Do lower income areas have more pedestrian casualties? *Accident Analysis and Prevention*.

- Paxton, P., Curran, P. J., Bollen, K. A., Kirby, J., and Chen, F. (2001). Monte Carlo experiments: Design and implementation. *Structural Equation Modeling*, 8(2):287–312.
- Ramsey, P. H. and Ramsey, P. P. (2009). Power and Type I errors for pairwise comparisons of means in the unequal variances case. *British Journal of Mathematical and Statistical Psychology*, 62:263–281.
- Sayegh, M. A., Castrucci, B. C., Lewis, K., and Hobbs-Lopez, A. (2010). Teen pregnancy in Texas: 2005 to 2015. *Maternal and Child Health Journal*, 14(1):94–101.
- Schönbrodt, F. D. and Perugini, M. (2013). At what sample size do correlations stabilize? *Journal of Research in Personality*, 47(5):609–612.
- Skrondal, A. (2000). Design and analysis of Monte Carlo experiments: Attacking the conventional wisdom. *Multivariate Behavioral Research*, 35(2):137—167.