# Assessment of the computational reproducibility of

# Fraley et al. (2022) N-Pact Factors

# Ian Hussey

# Summary

I was interested in conducting an N-Pact Factor analysis in a different domain, and therefore downloaded Fraley et al.’s (2022) code to attempt to reuse it. Before I did so, in order to check that I understood the code I assessed its computational reproducibility by comparing the results generated by the R script to those reported in the manuscript. No attempt was made to compare Fraley’s dataset to the results reported in the original articles (i.e., to re-extract any sample sizes or designs).

The authors’ substantive conclusions are reproducible from the their data and code. However, a subset of results were not computationally reproducible, and another subset of results were reproducible but contained errors (i.e., reproducible errors in the calculation of results, not their reporting).

# Source of code and data

Downloaded from <https://osf.io/rvbxp/> on 2022-1103.

# Results in manuscript that were not computationally reproducible

To calculate median N and k studies, the variables nPS2011, nPS2012, nPS2013 and nPS2014 are filtered for data6$Social\_PS==1 (e.g., line 214, 673), but no variable 'Social' is in the dataset. Table produced by code therefore has missing values where the manuscript’s table 1 does not. If this is changed to 'Social\_PS' then the tables are reproduced.

# Results in manuscript that were computationally reproducible but erroneous

Table 1 reports median sample sizes and K studies they were computed from. The row for PSPB reports median N and K studies values that are duplicated from Psych Science’s row. (e.g., line 299 in analyses\_original.R). This error is attributable to repetitive code that wasn't updated to the correct value between lines (i.e. a good old fashioned typo). My refactoring of the original code (see analyses\_refactored.Rmd and analyses\_refactored.html) uses a dplyr workflow that removes the possibly of this form of error.

Table 2 repeats this error when calculating power from (erroneous) median Ns for PSPB.

Table 3 repeats this error when calculating the FDR (I think, didn't deep dive this one as much).

# Results in manuscript that were computationally reproducible and correctly implemented in code but potentially misleading

Not an error in the manuscript but confusing in the code: the manuscript reports the False Discovery Rate (FP / [FP + TP]) and the code does indeed implement the calculation of this correctly, however this variable is labelled "fpr" in the code rather than “fdr” (i.e., False Positive Rate, which is instead FP / [FP + TN]).

# Stylistic choice

Rounding in the code was done via R's round(), which produces weird behaviour that most readers don’t expect. janitor::round\_half\_up() produces more intuitive behavior, but that's just my personal opinion. I only note this here because my refactoring of Fraley's code (see analyses\_refactored.Rmd and analyses\_refactored.html) uses janitor::round\_half\_up() and therefore produces medians that differ by 1 in about 15% of cases in Table 1.