PRDA: An R package for Prospective and Retrospective Design Analysis

Claudio Zandonella Callegher[1], **Giulia Bertoldo**[1], **Enrico Toffalini**[3], **Anna Vesely**[2], **Angela Andreella**[2], **Massimiliano Pastore**[1], and **Gianmarco Altoè**[1]

1 Department of Developmental Psychology and Socialisation, University of Padova, Padova, Italy **2** Department of Statistical Sciences, University of Padova, Padova, Italy **3** Department of General Psychology, University of Padova, Padova, Italy

Summary

Design Analysis was introduced by Gelman & Carlin (2014) as an extension of Power Analysis. Traditional power analysis has a narrow focus on statistical significance. Design analysis, instead, evaluates together with power levels also other inferential risks (i.e., Type M error and Type S error), to assess estimates uncertainty under hypothetical replications of a study.

Given an hypothetical value of effect size and study characteristics (i.e., sample size, statistical test directionality, significance level), *Type M error* (Magnitude, also known as Exaggeration Ratio) indicates the factor by which a statistically significant effect is on average exaggerated. *Type S error* (Sign), instead, indicates the probability of finding a statistically significant result in the opposite direction to the hypothetical effect.

Although Type M error and Type S error depend directly on power level, they underline valuable information regarding estimates uncertainty that would otherwise be overlooked. This enhances researchers awareness about the inferential risks related to their studies and helps them in the interpretation of their results. However, design analysis is rarely applied in real research settings also for the lack of dedicated software.

To know more about design analysis consider Gelman & Carlin (2014) and Lu et al. (2018). While, for an introduction to design analysis with examples in psychology see Altoè et al. (2020) and Bertoldo et al. (2020).

Statement of need

PRDA is an R package performing prospective or retrospective design analysis to evaluate inferential risks (i.e., power, Type M error, and Type S error) in a study considering Pearson's correlation between two variables or mean comparisons (one-sample, paired, two-sample, and Welch's *t*-test). *Prospective Design Analysis* is performed in the planning stage of a study to define the required sample size to obtain a given level of power. *Retrospective Design Analysis*, instead, is performed when the data have already been collected to evaluate the inferential risks associated with the study.

Another recent R package, retrodesign (Timm, 2019), allows conducting retrospective design analysis considering estimate of the unstandardized effect size (i.e., regression coefficient or mean difference) and standard error of the estimate. PRDA package, instead, considers standardized effect size (i.e., Pearson correlation coefficient or Cohen's *d*) and study sample

size. These are more commonly used in research fields such as Psychology or Social Science, and therefore are implemented in PRDA to facilitate researchers' reasoning about design analysis. PRDA, additionally, offers the possibility to conduct a prospective design analysis and to account for the uncertainty about the hypothetical value of effect size. In fact, hypothetical effect size can be defined as a single value according to previous results in the literature or experts indications, or by specifying a distribution of plausible values.

The package is available from GitHub (https://github.com/ClaudioZandonella/PRDA) and CRAN (https://CRAN.R-project.org/package=PRDA). Documentation about the package is available at https://claudiozandonella.github.io/PRDA/.

Examples

Imagine a study evaluating the relation a given personality trait (e.g., introversion) and math performance. Suppose that 20 participants were included in the study and results indicated a statistically significant correlation (e.g, $r = .55, p = .012$). The magnitude of the estimated correlation, however, is beyond what could be considered plausible in this field.

Retrospective design analysis

Suppose previous results in the literature indicate correlations in this area are more likely to be around $\rho = .25$. To evaluate the inferential risks associated with the study design, we can use the function `retrospective()`.

```r
library(PRDA)

set.seed(2020) # set seed to make results reproducible

retrospective(effect_size = .25, sample_n1 = 20, test_method = "pearson")
```

```
##
##  Design Analysis
##
## Hypothesized effect:  rho = 0.25
##
## Study characteristics:
##    test_method   sample_n1   sample_n2   alternative   sig_level   df
##    pearson       20          NULL        two_sided     0.05        18
##
## Inferential risks:
##    power   typeM   typeS
##    0.185   2.161   0.008
##
## Critical value(s): rho   =   ± 0.444
```

In the output, we have the summary information about the hypothesized population effect, the study characteristics, and the inferential risks. We obtained a statistical power of almost 20% that is associated with a Type M error of around 2.2 and a Type S error of 0.01. That means, statistical significant results are on average an overestimation of 120% of the hypothesized population effect and there is a 1% probability of obtaining a statistically significant result in the opposite direction. To know more about function arguments and examples see the function documentation and vignette.

Effect size distribution

Alternatively, if no precise information about hypothetical effect size is available, researchers could specify a distribution of values to account for their uncertainty. For example, they might define a normal distribution with mean of .25 and standard deviation of .1, truncated between .10 and 40.

```r
retrospective(effect_size = function(n) rnorm(n, .25, .1), sample_n1 = 20,
              test_method = "pearson", tl = .1, tu = .4, B = 1e3,
              display_message = FALSE)
```

```
## Truncation could require long computational time


##
##  Design Analysis
##
## Hypothesized effect:  rho ~ rnorm(n, 0.25, 0.1) [tl =  0.1 ; tu = 0.4 ]
##    n_effect   Min.    1st Qu.   Median   Mean    3rd Qu.   Max.
##    1000       0.101   0.197     0.25     0.252   0.308     0.4
##
## Study characteristics:
##    test_method    sample_n1    sample_n2    alternative    sig_level    df
##    pearson        20           NULL         two_sided      0.05         18
##
## Inferential risks:
##          Min.    1st Qu.   Median    Mean       3rd Qu.   Max.
## power    0.055   0.133     0.1880    0.203727   0.26600   0.449
## typeM    1.407   1.785     2.1645    2.347745   2.70075   5.263
## typeS    0.000   0.000     0.0060    0.017573   0.02300   0.246
##
## Critical value(s): rho   =   ± 0.444
```

Consequently this time we obtained a distribution of values for power, Type M error, and Type S error. Summary information are provided in the output.

Prospective design analysis

Given the previous results, researchers might consider planning a replication study to obtain more reliable results. The function `prospective()` can be used to compute the sample size needed to obtain a given level of power (e.g., power = 80%).

```r
prospective(effect_size = .25, power = .8, test_method = "pearson",
            display_message = FALSE)
```

```
##
##  Design Analysis
##
## Hypothesized effect:  rho = 0.25
##
## Study characteristics:
##    test_method    sample_n1    sample_n2    alternative    sig_level    df
##    pearson        122          NULL         two_sided      0.05         120
```

Callegher et al., (2021). PRDA: An R package for Prospective and Retrospective Design Analysis. *Journal of Open Source Software*, 6(58), 2810. https://doi.org/10.21105/joss.02810

```
##
## Inferential risks:
##    power   typeM   typeS
##    0.796   1.12    0
##
## Critical value(s): rho  =  ± 0.178
```

In the output, we have again the summary information about the hypothesized population effect, the study characteristics, and the inferential risks. To obtain a power of around 80% the required sample size is $n = 122$, the associated Type M error is around 1.10 and the Type S error is approximately 0. To know more about function arguments and examples see the function documentation and vignette.

In PRDA there are no implemented functions to obtain graphical representations of the results. However, it is easy to access all the results and use them to create the plots according to your own needs and preferences. See vignettes for an example.

References

Altoè, G., Bertoldo, G., Zandonella Callegher, C., Toffalini, E., Calcagnì, A., Finos, L., & Pastore, M. (2020). Enhancing Statistical Inference in Psychological Research via Prospective and Retrospective Design Analysis. *Frontiers in Psychology*, *10*. https://doi.org/10.3389/fpsyg.2019.02893

Bertoldo, G., Zandonella Callegher, C., & Altoè, G. (2020). *Designing Studies and Evaluating Research Results: Type M and Type S Errors for Pearson Correlation Coefficient* [Preprint]. PsyArXiv. https://doi.org/10.31234/osf.io/q9f86

Gelman, A., & Carlin, J. (2014). Beyond Power Calculations: Assessing Type S (Sign) and Type M (Magnitude) Errors. *Perspectives on Psychological Science*, *9*(6), 641–651. https://doi.org/10.1177/1745691614551642

Lu, J., Qiu, Y., & Deng, A. (2018). A note on Type S/M errors in hypothesis testing. *British Journal of Mathematical and Statistical Psychology*. https://doi.org/10.1111/bmsp.12132

Timm, A. (2019). *Retrodesign: Tools for type s (sign) and type m (magnitude) errors* [Manual]. https://github.com/andytimm/retrodesign