SMARTAR: An R package for designing and 1 analyzing Sequential Multiple Assignment 2 Randomized Trials

Review: By Pranav Pandit

Package Review

This review template is based on

- **Briefly describe any working relationship you have (had) with the package authors.**
- [x] As the reviewer I confirm that there are no conflicts of interest for me to review this work (If you are unsure whether you are in conflict, please speak to your editor *before* starting your review).

Instructions for authors:

Please complete following template as a part of package review. The package and manuscript with its documentation as it stands right now satisfies criteria that have been marked as tick.
Please complete the remaining criteria that have been marked as cross tick.

Documentation

The package includes all the following forms of documentation:

- [x] **A statement of need** clearly stating problems the software is designed to solve and its target audience in README
- [x] **Installation instructions:** for the development version of package and any non-standard dependencies in README
- ☒ **Vignette(s)** demonstrating major functionality that runs successfully locally
- ☒ **Function Documentation:** for all exported functions in R help
- ☒ **Examples** for all exported functions in R Help that run successfully locally
- ☒ **Community guidelines** including contribution guidelines in the README or CONTRIBUTING, and DESCRIPTION with `URL`, `BugReports` and `Maintainer` (which may be autogenerated via `Authors@R`).

> ***For packages co-submitting to JOSS***
>
> - [x] The package has an **obvious research application** according to JOSS's definition
>
> The package contains a `paper.md` matching JOSS's requirements with:
>
> - [x] **A short summary** describing the high-level functionality of the software
> - [x] **Authors:** A list of authors with their affiliations
> - [x] **A statement of need** clearly stating problems the software is designed to solve and its target audience.
> - [x] **References:** with DOIs for all those that have one (e.g. papers, datasets, software).

Functionality

- [x] **Installation:** Installation succeeds as documented.
- [&#x2612] **Functionality:** Any functional claims of the software been confirmed. Partial... there are few functions that do not perform as shown in the manuscript. Those tiny details are given in the missing functions list
- [NA] **Performance:** Any performance claims of the software been confirmed.
- [&#x2612] **Automated tests:** Unit tests cover essential functions of the package and a reasonable range of inputs and conditions. All tests pass on the local machine.
- [&#x2612] **Packaging guidelines**: The package conforms to the rOpenSci packaging guidelines

Missing functionalities for reproducibility and package utilization

- `help(SMARTAR)` does not produce help page.
- error in running `seqmeans(data=codiacs , family= "gaussian" , plot= "s" )` I would recommend including text or an example for how to further manipulate this graphs as per custom needs? Based on what I have seen these graphs are based solely on base R potting functions. Authors would greatly benefit if they could include examples related to custom editing the axis labels, colors options, marker options and other customizable plot functions.
- error in running `atsmeans (data=codiacs , family= 'gaissian' ,method='Gest' , digits =2, conf=TRUE,`

`alpha =0.05 , plot=TRUE)` this leads to following error
`Error in atsmeans(data = codiacs, family = "gaissian", method = "Gest", : unused argument (digits = 2)`
`Traceback:` -Missing explanation of `getncp` function. Mention of this function is missing in the manuscript completely.

Final approval (post-review)

- [ ] The author has responded to my review and made changes to my satisfaction. I recommend approving this package.

Estimated hours spent reviewing: 6 hours

- [x] Should the author(s) deem it appropriate, I agree to be acknowledged as a package reviewer ("rev" role) in the package DESCRIPTION file.

Review Comments

Review: SMARTAR: An R package for designing and 1 analyzing Sequential Multiple Assignment Randomized Trials

Authors present here an R package that they have developed for Sequential Multiple Assignment Randomized Trials (SMART). The overall manuscript is well developed, easy to read and the data is nicely represented. During this review I went through the rOpenSci's guidelines for reviewing a package too and checked if the manuscript fits in the review template provided by the rOpenSci community.

Here are my detailed comments. Within the introduction, authors explain in the SMART trial approach. They illustrate how the approach is essential in developing trails for adaptive treatments with a couple of simple figures. Following that authors describe the current milieu of R packages that can be used for analyses related to SMART and briefly describe how the new package is advantageous over the others.

Authors describe in detail all functions presented in the package and that are required for the analytical pipeline of SMART. Besides tiny missing details, authors have described the functioning of the package very well. I would say that a one important aspect of the case study is missing though that is interpretation of results and outputs. To address those, I would recommend authors to first start with the Sample Size function (as a first step of SMART planning) followed by data visualization and statistical comparison. At each step, authors can explain the functioning for the package with code and data. This should then include discussion regards to the results.

For example: in lines 307-311 authors say "In this case, conducting a global test among the 8 strategies 308 using the CODIACS trial data of 108 patients, it returned a test statistics Q = 36.03. Under 309 the null hypothesis, Q follows a centralized chi-squared distribution with degrees of freedom 310 v = 5. As a result, we obtained a P-value less than 0.05 and claimed over statistical significant 311 difference among the eight depression management strategies in the CODIACS trial." This should be followed by discussing the significance in terms of ATSs and their outcomes. I would recommend seeing Elith et al. 2008 A working guide to boosted regression trees. To follow the idea of explaining the biological interpretations of package developed.

Another comments I would include is to develop extensive documentation. Including various options for arguments. Authors provide some explanations in the text, but a centralized documentation is also essential as eventually users and developers would mostly refer to the package documentation. Similarly, help(SMARTAR) did not yield any documentation within my IDE while testing the package. There are additional small comments that I have noted while reviewing the package and they are attached below with the review.

Installation

In [1]:

```
install.packages("SMARTAR")
```

```
package 'SMARTAR' successfully unpacked and MD5 sums checked

The downloaded binary packages are in
 C:\Users\Falco\AppData\Local\Temp\RtmpkJLvvh\downloaded_packages
```

In [1]:

```
library(SMARTAR)
```

```
Warning message:
"package 'SMARTAR' was built under R version 3.6.3"
```

In [2]:

```
help(SMARTAR)
```

In [3]:

```
data(codiacs)
```

In [4]:

```
codiacs[1:5,]
```

A data.frame: 5 × 6

	ID	A1	O2	A2	Y	group
	<int>	<int>	<int>	<int>	<int>	<int>
1	1	1	1	1	12	8
2	2	0	0	0	10	1
3	3	1	0	1	7	6
4	4	1	0	1	2	6
5	5	1	1	1	8	8

In [5]:

```
lsf.str("package:SMARTAR")
```

```
atsmeans : function (data, family = "normal", method = "Gest", common = FALSE, conf = TRUE,
    alpha = 0.05, plot = FALSE)
getncp : function (df, alpha = 0.05, beta = 0.2, d = 1e-04, start = 5, dec = 2)
seqmeans : function (data, family = "normal", plot = "d")
smartest : function (data, family = "normal", method = "Gest", common = FALSE, alpha = 0.05,
    adjust = NULL, ntest = NULL)
smartsize : function (sim = NULL, delta = NULL, df = NULL, alpha = 0.05, beta = 0.2,
    global = TRUE, family = "normal", method = "Gest")
```

In [7]:

```
seqmeans(data=codiacs , family= "gaussian" , plot= "d" )#digits =2
```

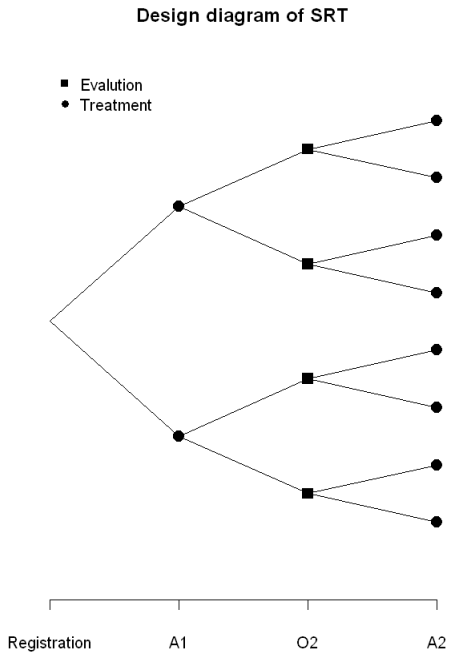
```
Each subject followed one of the below treatment sequences during the trial.

A treatment sequence is defined as a vector of values (A1,O2,A2).
```

A data.frame: 8 × 7

SEQ	A1	O2	A2	N	MEAN	VAR
<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<lgl>	<lgl>
1	0	0	0	25	NA	NA
2	0	0	1	2	NA	NA
3	0	1	0	24	NA	NA
4	0	1	1	5	NA	NA
5	1	0	0	5	NA	NA
6	1	0	1	19	NA	NA
7	1	1	0	2	NA	NA

SEQ	A1	O2	A2	N	MEAN	VAR
8	1	1	1	26	NA	NA



In [6]:

```
seqmeans(data=codiacs , family= "gaussian" , plot= "s" )#digits =2
```

```
Each subject followed one of the below treatment sequences during the trial.

A treatment sequence is defined as a vector of values (A1,O2,A2).
```

```
Error in title(main = ti, cex = 0.8): plot.new has not been called yet
Traceback:

1. seqmeans(data = codiacs, family = "gaussian", plot = "s")
2. seqplot(data = D, family = FA)
3. title(main = ti, cex = 0.8)
```

In [15]:

```
atsmeans(data=codiacs , family= 'gaissian' ,method='Gest' , conf=TRUE, alpha =0.05 , plot=TRUE)#digits = 2
```

```
$value: estimated strategy values (with confidence intervals)
$vmat: variance-covariance matrix of estimated strategy values


A strategy is defined as a vector of decision makings (d0;d00,d01) for 2 stages

d0 is the stage-1 decision making for A1
d00 is the stage-2 decision making for A2, conditioning on A1=d0 and O2=0
d01 is the stage-2 decision making for A2, conditioning on A1=d0 and O2=0
```

$value

A data.frame: 8 × 9

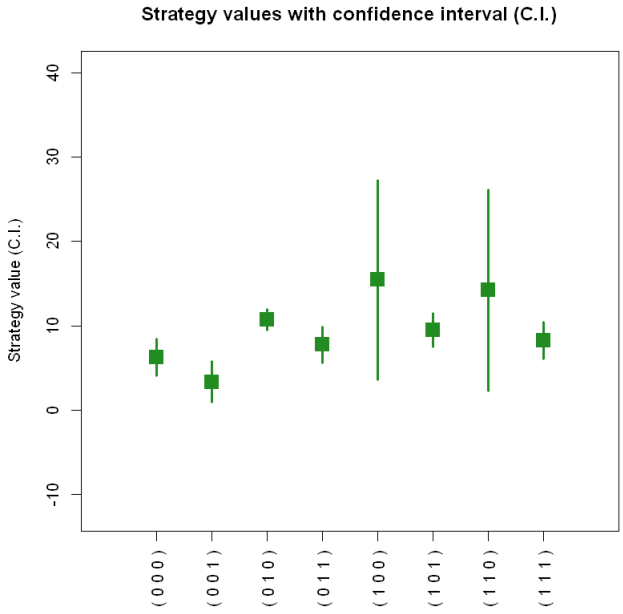
ATS	d0	d00	d01	N	value	se	lower	upper
<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
1	0	0	0	49	6.27	1.11	4.10	8.44

ATS	d0	d00	d01	N	value	se	lower	upper
<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
4	0	1	1	7	7.76	1.09	5.62	9.89
5	1	0	0	7	15.45	6.03	3.62	27.27
6	1	0	1	31	9.46	1.01	7.47	11.45
7	1	1	0	21	14.23	6.08	2.31	26.14
8	1	1	1	45	8.24	1.13	6.02	10.46

$vmat

A matrix: 8 × 8 of type dbl

1.2275	0.6347	0.3670	-0.2258	0.0000	0.0000	0.0000	0.0000
0.6347	1.5395	0.0065	0.9113	0.0000	0.0000	0.0000	0.0000
0.3670	0.0065	0.4098	0.0493	0.0000	0.0000	0.0000	0.0000
-0.2258	0.9113	0.0493	1.1863	0.0000	0.0000	0.0000	0.0000
0.0000	0.0000	0.0000	0.0000	36.4172	0.5800	36.2258	0.3886
0.0000	0.0000	0.0000	0.0000	0.5800	1.0302	0.2483	0.6985
0.0000	0.0000	0.0000	0.0000	36.2258	0.2483	36.9483	0.9708
0.0000	0.0000	0.0000	0.0000	0.3886	0.6985	0.9708	1.2806



In [17]:

```
smartest(data=codiacs, family="gaussian",  method="Gest", adjust= "Bon")#digits=2,
```

```
$Strategy provides the details of decision makings under strategy labels (ATS)
$Global.test assesses the null hypothesis of no difference across all the strategy values
$Pairwise.test compares all the pairs of strategies, of which the labels are shown in $Strategy

The P values should compare to the critical value adjusted for the Bonferroni correction
```

$Strategy

A matrix: 8 × 5 of type dbl

ATS	d0	d00	d10	N
1	0	0	0	49
2	0	0	1	30
3	0	1	0	26

ATS	d0	d00	d10	N
4	0	1	1	7
5	1	0	0	7
6	1	0	1	31
7	1	1	0	21
8	1	1	1	45

$Global.test

A data.frame: 1 × 5

size	nATS	df	chisq	Pvalue
<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
108	8	5	36.03	0

$Pairwise.comparisons

A data.frame: 56 × 6

label	diff	lower.CI	upper.CI	Z	Pvalue
<fct>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
1 vs. 2	2.94	-0.88	6.76	2.40	0.0163
1 vs. 3	-4.43	-7.39	-1.46	-4.66	0.0000
1 vs. 4	-1.49	-6.77	3.79	-0.88	0.3796
1 vs. 5	-9.18	-28.32	9.96	-1.50	0.1347
1 vs. 6	-3.19	-7.88	1.50	-2.12	0.0336
1 vs. 7	-7.96	-27.24	11.32	-1.29	0.1977
1 vs. 8	-1.97	-6.91	2.97	-1.25	0.2127
2 vs. 1	-2.94	-6.76	0.88	-2.40	0.0163
2 vs. 3	-7.36	-11.71	-3.02	-5.29	0.0000
2 vs. 4	-4.43	-7.39	-1.46	-4.66	0.0000
2 vs. 5	-12.12	-31.34	7.11	-1.97	0.0492
2 vs. 6	-6.13	-11.13	-1.13	-3.83	0.0001
2 vs. 7	-10.90	-30.25	8.46	-1.76	0.0790
2 vs. 8	-4.91	-10.15	0.33	-2.93	0.0034
3 vs. 1	4.43	1.46	7.39	4.66	0.0000
3 vs. 2	7.36	3.02	11.71	5.29	0.0000
3 vs. 4	2.94	-0.88	6.76	2.40	0.0163
3 vs. 5	-4.75	-23.69	14.18	-0.78	0.4336
3 vs. 6	1.23	-2.51	4.98	1.03	0.3041
3 vs. 7	-3.53	-22.60	15.54	-0.58	0.5633
3 vs. 8	2.45	-1.60	6.51	1.89	0.0592
4 vs. 1	1.49	-3.79	6.77	0.88	0.3796
4 vs. 2	4.43	1.46	7.39	4.66	0.0000

label	diff	lower.CI	upper.CI	t	Pvalue
<fct>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>
4 vs. 3	-2.99	-6.76	0.88	-2.42	0.0163
4 vs. 5	-7.69	-26.82	11.44	-1.25	0.2098
4 vs. 6	-1.71	-6.35	2.94	-1.15	0.2520
4 vs. 7	-6.47	-25.74	12.80	-1.05	0.2947
4 vs. 8	-0.49	-5.39	4.41	-0.31	0.7569
5 vs. 1	9.18	-9.96	28.32	1.50	0.1347
5 vs. 2	12.12	-7.11	31.34	1.97	0.0492
5 vs. 3	4.75	-14.18	23.69	0.78	0.4336
5 vs. 4	7.69	-11.44	26.82	1.25	0.2098
5 vs. 6	5.99	-12.81	24.78	0.99	0.3204
5 vs. 7	1.22	-1.76	4.20	1.28	0.2021
5 vs. 8	7.20	-11.75	26.16	1.19	0.2357
6 vs. 1	3.19	-1.50	7.88	2.12	0.0336
6 vs. 2	6.13	1.13	11.13	3.83	0.0001
6 vs. 3	-1.23	-4.98	2.51	-1.03	0.3041
6 vs. 4	1.71	-2.94	6.35	1.15	0.2520
6 vs. 5	-5.99	-24.78	12.81	-0.99	0.3204
6 vs. 7	-4.77	-23.87	14.34	-0.78	0.4363
6 vs. 8	1.22	-1.76	4.20	1.28	0.2021
7 vs. 1	7.96	-11.32	27.24	1.29	0.1977
7 vs. 2	10.90	-8.46	30.25	1.76	0.0790
7 vs. 3	3.53	-15.54	22.60	0.58	0.5633
7 vs. 4	6.47	-12.80	25.74	1.05	0.2947
7 vs. 5	-1.22	-4.20	1.76	-1.28	0.2021
7 vs. 6	4.77	-14.34	23.87	0.78	0.4363
7 vs. 8	5.99	-12.81	24.78	0.99	0.3204
8 vs. 1	1.97	-2.97	6.91	1.25	0.2127
8 vs. 2	4.91	-0.33	10.15	2.93	0.0034
8 vs. 3	-2.45	-6.51	1.60	-1.89	0.0592
8 vs. 4	0.49	-4.41	5.39	0.31	0.7569
8 vs. 5	-7.20	-26.16	11.75	-1.19	0.2357
8 vs. 6	-1.22	-4.20	1.76	-1.28	0.2021
8 vs. 7	-5.99	-24.78	12.81	-0.99	0.3204

label	diff	lower.CI	upper.CI	Z	Pvalue