Paper PO-13

A Randomization Test SAS/IML® Program for Making Treatment Effects Inferences for Extensions and Variations of ABAB Single-Case Experimental Designs

Patricia Rodríguez de Gil, University of South Florida, Tampa, FL
John M. Ferron, University of South Florida, Tampa, FL

ABSTRACT

While the evaluation of intervention effects in single-case research has relied on visual inspection of the data (Kazdin, 1980), the description of graphical forms are not considered an adequate substitute for statistical tests (Edgington, 1980). Moreover, there are cases when graphical displays of data tend to be quite ambiguous and treatment effects are not easily appreciated (Ferron & Sentovich, 2002); in these cases, inferential statistics are often necessary to determine if a treatment effect exists. Randomization tests are considered valid statistical tests for determining the presence of a treatment effect in single-case experimental data (Edgington, 1980). In addition, significance tests lead to a more informed and reflective statistical analysis (Thompson & Snyder, 1997). Although the statistical validity of randomization tests has been established, randomization tests for single-case data are not incorporated into readily available statistical software like SAS® and SPSS, making it difficult for researchers to implement randomization tests into their statistical analysis of data. The example provided for Onghena (1992) was used to illustrate a worked example of a randomization test where the use of random assignment of treatment to treatment times and the incorporation of randomization into single-case reversal designs is explained and applied to statistical testing. SAS/IML code for randomization tests for extensions and variations of ABAB single-case experimental designs is provided and discussed.

Keywords: SAS/IML, RANDOMIZATION, RANDOMIZATION TEST, TREATMENT EFFECTS, ABA, ABAB, ABABA, SINGLE-CASE, SINGLE-SUBJECT

INTRODUCTION

Traditionally, the evaluation of intervention effects in single-case research has relied on visual inspection of the data (Kazdin, 1980); Parsonson and Baer (1992) for instance, argued that inferential statistics are an unnecessary addition to visual inspections of the data. However, as Ferron and Sentovich (2002) asserted, there are cases where graphical displays of data tend to be quite ambiguous and treatment effects are not easily appreciated. Barlow and Hayes (1979) stated that because considerable intersubject variability exists (e.g., some participants change and some do not), inferential statistics are often necessary to determine if a treatment effect exists. The need for "good research," that efficiently determines the effectiveness of a treatment, and that considers statistical tests as a method for providing additional information regarding the presence of a treatment effect in single-case designs has been stressed by researchers (Edgington, 1980) and called upon by academic journals as well (Edgington, 1985). Randomization tests are one such method for determining significance of treatment effects in single-case experimental designs (Edgington, 1980).

RANDOMIZATION TESTS

What are randomization tests? Randomization tests, based on the assumption of random assignment (Edgington, 1980), are statistical, nonparametric or "distribution-free" (Edgington, 1985, p. 241) procedures for determining the statistical significance of experiments that can be applied to experimental data when the assumptions required for the application of parametric tests are not tenable (Edgington, 1980; Edgington & Khuller, 1992), as it is the case of single-case experiments.

One basic design in single-case analysis is that of ABAB. Kazdin (1982) described this experimental design as one where measurements for a single unit are taken under two different alternating conditions, in four phases and with several measurements in each phase. Using Campbell and Stanley's (1963) notation, ABAB experimental designs can be represented using the following diagram:

$$OO \cdots O\ X\ O\ O\ \cdots O\ X^{-}\ OO \cdots OXOO$$

where *X* represents the exposure to an experimental condition or event (the effects of which are to be measured) and *O* will refer to some process of observation or measurement. The X's and O's in a given row are applied to the same specific study participant. The left to right dimensions indicates the temporal order. In other words, the first phase is the baseline phase A_1 whose repeated measures are taken under controlled conditions; the second phase constitutes

the intervention phase B_1, where measures are also taken under controlled conditions. The third phase, A_2 is the withdrawal condition. Then, the last phase or phase B_2 is the second experimental intervention (Onghena, 1992). See Figure 1 below.

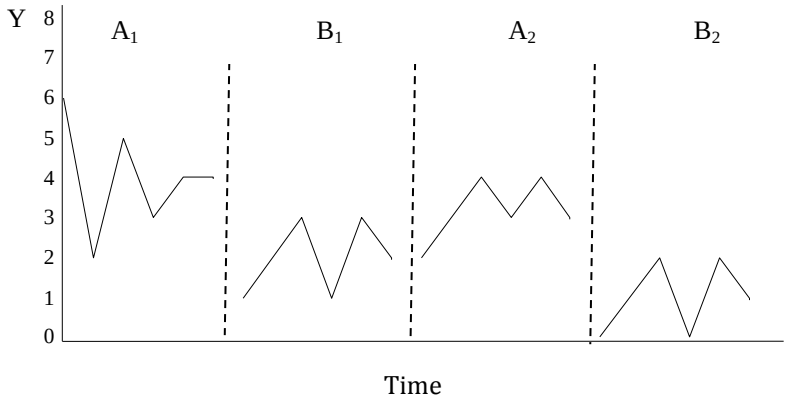

Figure 1. Basic ABAB Single-Subject Analysis

As it can be observed in Figure 1 above, because of the non-independence nature of the measurements in the ABAB single-case analysis, parametric statistical tests cannot by applied to such designs. Thus, we need to resort to statistical tests that do not require adherence to the independence assumption, such as randomization tests**.**

GENERAL NULL HYPOTHESIS APPLICABLE TO A RANDOMIZATION TEST IN SINGLE CASE EXPERIMENTS

Randomization tests compare an obtained test statistic to a randomization distribution. According to Edgington (1980), the null hypothesis applicable to a randomization test for any single-case experiment is that the sequence of measurements is the same as it would have been for any of the other possible assignments of treatment times (or intervals of time) to treatments.

Statistical significance by the randomization test procedure is determined by finding that the proportion of possible assignments that give test statistics as or more extreme than the obtained test statistic.

The randomization distribution is also based on assuming the null hypothesis is true, but the distribution of possible results is based on rearranging the data to consider the possible random assignments.

Assuming that the null hypothesis is true, that there is no differential effectiveness of the treatments, randomization tests determine the significance of experimental results by comparing the observed test statistic such as a *t*-ratio (difference between means) to a distribution of values obtained by dividing up the data in all possible ways consistent with the random assignment procedure and computing a test statistic value for each division.

The obtained *p*-value (level of significance reached) is the proportion of the data divisions with a test statistic equal to or greater than the observed *t*- value (Edgington, 1985).

PURPOSE

The purpose of this paper is to present an example data seta that shows the use of randomization tests for making treatment effects inferences for extensions and variations of ABAB single-case design and to provide SAS/IML programs that can be easily modified to carry out sound statistical analysis of single-case data. Advantages for using randomization tests in single-case analysis are also discussed.

METHOD

Assuming that previously to performing the randomization test, 1) the alternative null hypothesis has been selected, 2) the specific level of significance as well as 3) the number of measurement times, *N*, has been specified, 4) a test statistic has been selected, 5) data has been collected, and 6) computations of the test statistic from the obtained data have been performed (*T* statistic), a valid randomization test can be then performed (Onghena, 1992).

The example provided by Onghena (1992) is used to illustrate the underlying principle of the randomization test procedure, that of the random assignment of treatment to treatment times, and the incorporation of randomization into

an ABAB design [baseline (A_1), intervention (B_1), baseline (A_2), and intervention (B_2)], where a triplet of intervention points with a minimum phase length can be chosen randomly. That is, for an ABAB design, 1) time of the first intervention (T_1), 2) the time of the withdrawal (T_2), and 3) the time of the second intervention (T_3).

This random selection of the three points of change, under the constrain that the minimum phase length n = 4, will ensure that we will not have too few measurement times for one of the phases (Onghena, 1992). The alternative hypothesis for the randomization test was determined to be directional or one-tailed, and the level of significance was set to α = .05 and N = 24. Equally important at this time is the determination of the three points of change (k = 3).

DATA SET

The example data set for this randomization test demonstration is part of the hypothetical data of a randomization test used by Onghena (1992) and it is shown in Table 1 below, where the mean values of each phase are 4.0, 2.0, 3.0, and 1.0.

	Treatment Time																							
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24
Phase	A	A	A	A	A	A	B	B	B	B	B	B	A	A	A	A	A	A	B	B	B	B	B	B
Score	6	2	5	3	4	4	1	2	3	1	3	2	2	3	4	2	4	3	0	1	2	0	2	1

Table 1. Hypothetical Data (*N*=24, *k*=3)

$A_1 = 6 + 2 + 5 + 3 + 4 + 4 = 24/6 = 4$

$B_1 = 1 + 2 + 3 + 1 + 3 + 2 = 12/6 = 2$

$A_2 = 2 + 3 + 4 + 2 + 4 + 3 = 18/6 = 3$

$B_2 = 0 + 1 + 2 + 0 + 2 + 1 = \ 6/6 = 1$

Observed test statistic $T = (\bar{A}_1 + \bar{A}_2) - (\bar{B}_1 + \bar{B}_2)$

$$T = (4 + 3) \ - (2 + 1) = 4$$

RANDOMIZATION

For the data presented, once it has been determined that n = 4 (measurements in each phase), the formula to calculate the number of possible intervention triplets that met these constrains, where N = total measurement times, k = points of change, and n = minimum measurement times in each phase is:

$$\frac{[N - n(\mathrm{k} + 1) + \ k]!}{[N - n\ (\mathrm{k} + 10)] * k!}$$

Considering the data provided in Table 1 where N = 24, k = 3, and n = 4

$$\frac{[24 - 4(3 + 1) + \ 3]!}{[24 - 4(3 + 1)]! \ * \ 3!} = \frac{11}{8! * 3!} = 165$$

Note that for a study with 24 observations and a minimum phase length set to 4, 165 possible data combinations

were involved. The obtained test statistic (*T*), will be then compared to a distribution of values by dividing up the data in all possible ways consistent with the random assignment procedure and computing a test statistic value for each division. The obtained *p*-value (level of significance reached) is the proportion of the data divisions with a test statistic equal or greater than the obtained value (Edgington, 1985).

RESULTS

An exhaustive randomization test for an ABAB design was conducted, where all possible test statistics were calculated. There were 165 possible assignments of treatments to treatment times. The obtained test statistic (*T* = 4) was compared to the randomization distribution that was constructed considering all possible assignments. The probability of obtaining a test statistic as large or larger than 4 was found to be statistically significant (*p* = 0. 024).

SAS/IML Program Description for a Randomization Test for an ABAB Reversal Design

As mentioned before, although the statistical validity of randomization tests has been established, randomization tests for single-case data are not incorporated into readily available statistical software like SAS and SPSS, making it difficult for researchers to implement randomization tests into their statistical analysis of data. In this section, we first introduce the example data set, which will be used to explain the SAS program that was created to make inferences on the power of each of the three types of reversal designs (ABA, ABAB, and ABABA) using a randomization test.

The test was run using one participant. Also, since the minimum number of possible assignments needed to reach statistical significance at .05 is at least 20 (e.g., 1/20 = .05), each run was done using the 24 observations in the hypothetical data set. The study participant was randomly assigned to different series lengths. For the 24 observations, the observations were distributed as depicted below for an ABAB reversal design.

- N=24 k=3 n=4
- C=(N – n(k+1)+k)! / (N – n(k+1))!k! = 165
- N = Number of observations
- k = number of points of change
- n = minimum observations per phase
- C= number of possible random assignments

Analysis of the Sample using SAS/IML

In our example, we are interested in using a randomization test to perform a test of statistical significance in single-case data by dividing the example data set in all possible ways consistent with the random assignment procedure. We will use the following specification for an ABAB design, *N* = 24, *k* = 3, n = *4* (SAS code for ABA and ABABA designs are provided at the end of this paper).

The fundamental element in SAS/IML is the matrix, a two-dimensional row x column array of numeric values. Thus, we will start by inputting values in the X matrix, the scores from the 24 observation of the example data set.

```
proc iml;

x={6, 2, 5, 3, 4, 4,
   1, 2, 3, 1, 3, 2,
   2, 3, 4, 2, 4, 3,
   0, 1, 2, 0, 2, 1};
```

Next, the observed scores are grouped to create 4 sets (for the ABAB single-case design)

```
a1=x[1:6];      b1=x[7:12];
a2=x[13:18];    b2=x[19:24];
```

Now, the observed average score (“obs”) is calculated:

```
obs=(sum(a1)/6 + sum(a2)/6 )/2 -
    (sum(b1)/6 + sum(b2)/6 )/2;
```

The obtained test statistic such as a t-ratio (difference between means) will be then compared to a distribution of values by dividing up the data in all possible ways consistent with the random assignment procedure and computing a test statistic value for each division. The obtained *p*-value (level of significance reached) is the proportion of the

data divisions with a test statistic equal or greater than the obtained value (Edgington, 1985).

```
do j=(mpl+1) to (nn-(mpl*3-1));
    do k=(j+mpl) to (nn-(mpl*2-1));
        do m=(k+mpl) to (nn-(mpl-1));

      rncount=rncount+1;

      a1=x[1:j-1]; b1=x[j:k-1];
      a2=x[k:m-1]; b2=x[m:nn];

      teststat= (sum(a1)/nrow(a1) + sum(a2)/nrow(a2))/ 2 -
                (sum(b1)/nrow(b1) + sum(b2)/nrow(b2))/ 2;
```

The score for each of the 165 possible combinations will be calculated using 3 do loops (3 points of change for ABAB design) and saved in “teststat” variable.

```
if teststat > = obs then count = count + 1;
    end;
  end;
end;
```

If the calculated score is greater than the obtained test statistic, add 1 to counter.

```
pvalue=count/rncount;
print pvalue count rncount obs;
quit;
```

This last part of the program calculates and prints the results; *p-value* (level of significance reached) is the proportion of the data divisions with a test statistic equal or greater than the obtained value (Edgington, 1985).

ADVANTAGES OF IMPLEMENTING RANDOMIZATION TESTS IN REVERSAL DESIGNS

Among the many advantages for the use of randomization tests for making treatment effects inferences for extensions and variations of ABAB Single-Case Experimental Designs, we can mention the increase in objectivity that is provided by such tests. While the analysis of graphical displays of the single-case data can provide valuable information about the behavior being observed, randomization tests provide an exact statistical test. The nature of single-case data do not allow the parametric assumptions (e.g., normality of distribution, independence of observations). In such case, randomization tests are distribution-free tests that allow us to perform a statistical test, preserving without alteration, the variability of the original data

In using the reversal technique, the experimenter is attempting to show that an analysis of the behavior is at hand, that whenever he applies certain condition, the behavior is produced, and whenever the condition is removed, the behavior is lost. Although the randomization test does not alter the ABAB structure, it alters the lengths of the phases and that could raise concern since stabilization of the data is an important criterion in the decision to change conditions.

Because the validity of randomization tests depends on the random assignment of treatment times to treatments (Edgington, 1985), this test is at odds with response-guided experimentation and the utility of randomization tests has been questioned (Kazdin, 1980). However, there are ways to integrate randomization and response-guided methodologies (Ferron, 1994).

SAS Programs for ABAB, ABA, and ABABA Single-Case

```
/*-------------------------------------------------*
 * Program 1. ABAB N=24 k=3 n=4      *
 *-------------------------------------------------*/

PROC IML;

x={6, 2, 5, 3, 4, 4,
   1, 2, 3, 1, 3, 2,
   2, 3, 4, 2, 4, 3,
   0, 1, 2, 0, 2, 1};

a1=x[1:6];      b1=x[7:12];
a2=x[13:18];    b2=x[19:24];

obs=(sum(a1)/6 + sum(a2)/6 ) / 2 -
    (sum(b1)/6 + sum(b2)/6 ) / 2;

mpl=4;
rncount=0;
count=0;
nn=nrow(x);

do j=(mpl+1) to (nn-(mpl*3-1));
    do k=(j+mpl) to (nn-(mpl*2-1));
        do m=(k+mpl) to (nn-(mpl-1));
      rncount=rncount+1;
      a1=x[1:j-1]; b1=x[j:k-1];
      a2=x[k:m-1]; b2=x[m:nn];

      teststat= ( sum(a1)/nrow(a1) + sum(a2)/nrow(a2) ) / 2 -
                ( sum(b1)/nrow(b1) + sum(b2)/nrow(b2) ) / 2;

      if teststat >= obs then count = count + 1;
    end;
  end;
end;

pvalue=count/rncount;
print pvalue count rncount obs;
quit;

/*-------------------------------------------------*
* Program 2.  ABA   N=24 k=2 n=6 *
*-------------------------------------------------*/
PROC IML;
x={ 1, 1, 1, 1, 1, 1, 1, 1,
    0, 0, 0, 0, 0, 0, 0, 0,
    1, 1, 1, 1, 1, 1, 1, 1};

a1=x[1:8];
b1=x[9:16];
a2=x[17:24];

obs = (sum(a1)/8 + sum(a2)/8 ) / 2 -
      (sum(b1)/8 );

print a1 a2 b1 obs;

mpl = 6;            /* Minimum phase length */
nn = nrow(x);       /* total observations */
rncount = 0;
count = 0;

do j = (1 + mpl) to (nn - (mpl * 2 - 1));

print 'starting j loop ' j;

     do k = (j + mpl) to (nn - (mpl - 1));
```

```
rncount = rncount + 1;

     a1 = x[1 : j-1];
     b1 = x[j : k-1];
     a2 = x[k : nn];

        teststat = (sum(a1)/nrow(a1) + sum(a2)/nrow(a2) ) / 2 -
                   (sum(b1) / nrow(b1));

     if teststat >= obs then count = count + 1;

     print rncount count j k a1 b1 a2 teststat;

    end;
end;
pvalue = count / rncount;
print pvalue count rncount obs;
quit;
/*--------------------------------------*
/* Program 3. ABABA  N=24 k=4 n=4       *
*--------------------------------------*/
PROC IML;

x={6, 2, 5, 3,
   4, 4, 1, 2,
   3, 1, 3, 2,
   2, 3, 4, 2,
   4, 3, 0, 1, 2, 0, 2, 1};

a1=x[1:4];      b1=x[5:8];
a2=x[9:12];     b2=x[13:16];
a3=x[17:24];

obs=( (sum(a1)/4) + (sum(a2)/4) + (sum(a3)/8) ) / 3  -
    ( (sum(b1)/4) + (sum(b2)/4) ) / 2;

print a1 a2 a3 b1 b2 obs;

mpl=4;
rncount=0;
count=0;
nn=nrow(x);

do i=(1+mpl) to (nn - (mpl*4-1));

  print 'starting i loop ' i;

    do j=(mpl+1) to (nn - (mpl*3-1));
       do k=(j+mpl) to (nn - (mpl*2-1));
          do m=(k+mpl) to (nn - (mpl-1));

        rncount=rncount+1;

        a1=x[1 : i-1];       b1=x[i : j-1];
        a2=x[j : k-1];       b2=x[k : m-1];
        a3=x[m : nn];

        teststat=( ( sum(a1)/nrow(a1) + sum(a2)/nrow(a2) + sum(a3)/nrow(a3) ) / 3 ) -
                 ( ( sum(b1)/nrow(b1) + sum(b2)/nrow(b2) ) / 2 );

        if teststat>=obs then count=count + 1;

        print rncount count i j k m a1 a2 a3 b1 b2 teststat;

      end;
    end;
  end;
end;
pvalue=count/rncount;
print pvalue count rncount obs;
quit;
```

CONCLUSION

Modern computers have made procedures such as randomization tests "a more viable alternative to using classical parametrical methodologies" (Mackenzie & Manly, 2001, p. 292). Randomization tests often offer an alternative approach that can be implemented in a relatively straightforward manner. Statistical evaluation can help identify reliable effects that can be pursued and strengthened in further research (Kazdin, 1980, p. 254).

Although the use of randomization tests for determining the statistical significance of treatment effects in single-case analysis is still in debate (Ferron & Onghena, 1996), they provide valid procedures about the efficacy of treatment effects in single-case designs and ready access to easy procedures for implementing them should encourage researchers to use them to support their findings (Busk and Marascullo, 1992; Edgington, 1992).

REFERENCES

Baer, D. M., Wolf, M. M., & Risley, T. R. (1968). Some current dimensions of applied behavioral analysis. *Journal of Applied Behavioral Analysis, 1,* 91-97.

Barlow, D. H., & Hayes, S. C. (1979). Alternative treatments design: One strategy for comparing the effects of two *treatments in a single subject. Journal of Applied Behaviors Analysis, 12, 199-210.*

Busk, P. L., & Marascullo, L. A. (1992). Statistical Analysis is single-case research: Issues, procedures, and recommendations, with application to multiple behaviors. In T. R. Kratochwill & F. R. Levin (Eds.), *Single-Case Research Design and Analysis: New Directions for Psychology and Education, pp. 159-185.* Hillsdale, NJ: Erlbaum.

Edgington, E. S. (1980). Validity of randomization tests for one-subject experiments. *Journal of Educational Statistics, 5,* 235-251.

Edgington, E. S. (1985). Random assignment and experimental research. *Educational Administration Quarterly,21*(3), 235-246.

Ferron, J., & Sentovich, C. (2002). Statistical power of randomization tests used with multiple-baseline designs. *The Journal of Experimental Education, 70*(2), 165-178.

Ferron, J. & Ware, W. (1994). Using randomization tests with responsive single-case designs. Behavior Research and Therapy, 32, 787-791.

Kazdin, A. E. (1980). Obstacles in using randomization tests in single-case experimentation. *Journal of Educational Statistics, 5,* 253-260.

Onghena, P. (1992). Randomization test for extension and variations of ABAB single-case experimental design: A rejoinder. *Behavioral Assessment, 14,* 153-171.

Parsonson, , B. S., & Baer, D. M. (1992). The graphic analysis of data. In A. Polling & R. W. Fuqua (Eds.), *Research Methods in Applied Behavior Analysis, pp. 157-186.* NY: Plenum Publishing Corporation

Romano, J. P. (1990). On the behavior of randomization tests without a group invariance assumption. *Journal of the American Statistical Association, 85*, 686-692.

SAS Institute, Inc. (2002-2003). SAS version 9.2

Thompson, B., Snyder, P.A. (1997). Statistical significance practices in The Journal of Experimental Education. *The Journal of Experimental Education, 66*(1), 75-83.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Patricia Rodríguez de Gil
University of South Florida
4202 E. Fowler Ave. EDU 105
Tampa, FL 33620
E-mail: prodrig6@usf.edu