An OOM analog to the a priori procedure

© *James W. Grice, Ph.D.*
Oklahoma State University
Updated: June 8th, 2022

As an alternative to null hypothesis significance testing (i.e., the traditional *p*-value), David Trafimow developed the *a priori procedure* (APP) as a sensible and superior method for drawing inferences to population parameters from sample statistics. Perhaps the most prominent and pragmatic capability of the APP is the determination of sample size prior to conducting an empirical study. In his 2019 paper, *A Frequentist Alternative to Significance Testing, p-Values, and Confidence Intervals* , Trafimow (p. 6) provided the following instructive example:

> Trafimow (2017) provided an accessible proof of Equation (1) below:
>
> $$n = \left(\frac{z_C}{f}\right)^2, \quad (1)$$
>
> where
>
> - f is the fraction of a standard deviation the researcher defines as sufficiently "close,"
> - Z_C is the z-score that corresponds to the desired probability of being close, and
> - n is the minimum sample size necessary to meet specifications for closeness and confidence.
>
> For example, suppose the researcher wishes to have a 95% probability of obtaining a sample mean that is within a quarter of a standard deviation of the population mean. Because the z-score that corresponds to 95% confidence is approximately 1.96, Equation (1) can be solved as follows: $n = \left(\frac{1.96}{0.25}\right)^2 = 61.47 \approx 62$. In other words, the researcher will need to recruit 62 participants to meet specifications for closeness and confidence.[9]

Without the use of equations the logic of the APP can be followed when analyzing data in the OOM software. The application of the APP logic in OOM revolves around the randomization test. Recall that in OOM observations or deep structures can be randomized for any given analysis to create a distribution of random-data Percent Correct Classifications indices (PCCs). The observed PCC for the actual observations can then be compared to these random-data PCCs to conclude if the observed pattern is best explained as the product of physical chance (loosely similar to declaring a result as "statistically significant" using null hypothesis significance testing). When attempting to judge against physical chance in this manner, it is desirable for the distribution of random-data PCCs to possess low variability; in other words, it is desirable for the PCCs in the distribution to be tightly packed around their mean. For example, consider two distributions with equal means of 1000 PCCs from a randomization test:

A)

```
  0.00 : 10.00 : [ 40]**
 11.00 : 20.00 : [ 60]***
 21.00 : 30.00 : [ 80]****
 31.00 : 40.00 : [120]*****
 41.00 : 50.00 : [200]**********
 51.00 : 60.00 : [200]**********
 61.00 : 70.00 : [120]*****
 71.00 : 80.00 : [ 80]****
 81.00 : 90.00 : [ 60]***
 91.00 : 100.0 : [ 40]**
```

B)

```
  0.00 : 10.00 : [  0]
 11.00 : 20.00 : [  0]
 21.00 : 30.00 : [ 10]*
 31.00 : 40.00 : [ 50]***
 41.00 : 50.00 : [440]**********************
 51.00 : 60.00 : [440]**********************
 61.00 : 70.00 : [ 50]***
 71.00 : 80.00 : [ 10]*
 81.00 : 90.00 : [  0]
 91.00 : 100.0 : [  0]
```

The distribution on the left (A) demonstrates greater variability, as low and high PCCs are more frequent than in the distribution on the right (B). Image a researcher who observes a PCC index equal to 81% for the actual data. Such a value never appears in the distribution on the right (B), indicating that it is not easily produced by randomizing the data; thus, the researcher can choose against physical chance as a plausible explanation of the data pattern underlying the 81% PCC. By comparison, the distribution on the left (A) shows values of at least 81% are somewhat common; physical chance therefore remains as a plausible explanation for the data pattern.

The desired low variability in the distribution of random-data PCCs is the "closeness" referred to by Trafimow above as f in Equation 1. In OOM the Z_c value in Equation 1 is also applied to these PCCs, and it represents the proportion of values desired to be close to the mean of the random-data PCCs. In the spirit of Trafimow's example above, suppose a researcher plans to use a particular analysis in the OOM software and expects a PCC value equal to 50% if physical chance is the best explanation of the data. Based on theory, the researcher hopes for a PCC much higher, say, at least 85%. Suppose further the researcher wishes to collect enough data so that when the randomization test is conducted ~95% of the random-data PCCs will fall within 10 percentage points of 50%; that is, ~95% of the random-data PCCs will range in value from 40% to 60%. In this example 10% is the *closeness value*, and 95% is the desired *closeness proportion* of PCCs within the closeness range. These values are not to be considered as the f and Z_c values, per se, but only as analogs. In OOM the researcher does not use the APP equations, but instead works in trial-and-error fashion to arrive at the needed sample size once the desired closeness value and proportion have been decided upon.

The logic behind the APP can be packaged as a series of steps to follow when conducting analyses in most procedures in the OOM software. These steps are presented here, and then an example analysis is presented that demonstrates how the steps are executed in order to determine the sample size of a planned study.

Steps:

1. An integrated model is created to explicate the structures and processes of the natural system under investigation. As a consequence of the model, an hypothesis is put forth, a study is designed, and a particular OOM analysis is decided upon.
2. The *closeness value* and desired *closeness proportion* are chosen. There is no need to predict or estimate the expected mean of the PCCs from the randomization test. Only the closeness value and closeness proportion are needed. Suggested closeness values are 1%, 5%, or 10%. Suggested proportions are 90%, 95%, or 99%. These values are suggested solely for the sake of their familiarity to modern researchers. Other values can be freely and arbitrarily chosen.
3. An initial guestimate is made with regard to the needed sample size. "Guestimate" is used here because this process is not an exact science (nor need it be one). Suggested starting sample sizes are 50, 100, or 200 cases, but much will depend upon the analysis technique, number of orderings being analyzed, and the number of units of observation comprising the orderings.
4. A data set is created based on the starting sample size. The data set can be created in one of three ways:
    a. *Manually*. The data are entered directly into the *Data Edit* window of the OOM software. The data can be created to possess particular distributional properties, or the data can be created haphazardly.

b. *Generate Random Observations*. This option in the OOM software can be used to generate any number of random observations. Real or integer values can be created with approximate normal or uniform distributions.
c. *Generate Observations*. Working with integer or binned real values, a large number of observations that match an a priori distribution can be generated.

5. The selected OOM analysis is conducted using the generated data set. The *Randomization Test* option is selected with the desired randomization method, and the *Save Randomized Results* option is selected to save the generated PCCs into a new data set. Any number of randomization trials can be determined. Suggested numbers of trials are 1000, 5000, or 10000.
6. The mean of the saved, random-data PCCs is obtained from the *Descriptive Statistics* procedure in the OOM software.
7. Based on the mean and the closeness value, a range of PCCs is created, and the number of random-data PCCs included in the range are tallied. This work is done in the *Define Orderings* window of the OOM software. It is also important here to evaluate any a priori expectations regarding the PCC. If for instance, a researcher considers an observed PCC of 65% to offer compelling evidence for the posited theory, and the mean of random-data PCCs is also 65%, then the theory, methods, or analysis strategy is in need of rethinking.
8. If the proportion of random-data PCCs in the closeness range is approximately equal to the closeness proportion, then the current sample size is used for data collection. If the proportion of random-data PCCs in the closeness range is substantively less than the closeness proportion, then a larger sample size needs to be tested. If the proportion of random-data PCCs in the closeness range is substantively greater than the closeness proportion, then a smaller sample can be tested. In these latter two instances, a new guestimate for the sample size is selected by the researcher, and then Steps 4 through 8 are repeated until the proportion of random-data PCCs in the closeness range is approximately equal to the closeness proportion.

Example #1

Consider the recent study by Sawaoka and Monin (2018, Study 1; data available at https://osf.io/qtsx2/) in which participants viewed an offensive media post and then read fictitious comments that expressed moral outrage at the post. Participants were divided into *nonviral* and *viral* groups; the nonviral group read 2 comments, and the viral group read 10. Participants were then asked to evaluate one of the commenters (selected by the authors) for the following characteristics: "in the wrong," "a bully," "praiseworthy," and "a good person" (response scale from 1 to 7). The four items were keyed so that the scale was in the same direction for all of them, and responses were averaged to form the primary dependent variable for the study; high scores indicated a negative evaluation. Sawaoka and Monin theorized that the commenter evaluated in the viral, 10-comment condition would be seen as someone seeking social validation rather than as someone expressing authentic moral outrage. Consequently, they predicted that the evaluations would be more negative in the viral condition compared with the nonviral condition; i.e., *viral* > *nonviral*. Assuming Sawaoka and Monin used observation oriented modeling, the steps above could have been used to plan the sample size.

Step 1. Choose the analysis procedure in the OOM software. As Sawaoka and Monin are predicting an ordinal difference on ratings for the two groups (*viral* > *nonviral*), an *Ordinal Analysis – Crossed Orderings* procedure can be used. Sawaoka and Monin used a between-

subjects ANOVA to analyze their data, and they could have analyzed the simple two-group comparison with an independent samples *t*-test.

Step 2. Choose the closeness value and desired closeness proportion. Suppose Sawaoka and Monin chose 5% as the closeness value and 95% as the closeness proportion. In other words, their expectation is that the randomization test will yield a distribution of random-data PCCs with a particular mean. Approximately 95% of the random-data PCCs will fall within ±5% of this mean. For example, if the mean is 50%, then ~95% of the random-data PCCs will fall within the 45% to 55% range.

Step 3. Choose a starting sample size. Suppose Sawaoka and Monin chose 50 persons per group, for a total sample size of 100 persons.

Step 4. Create a data set based on the starting sample size. The *Generate Random Observations* procedure in the OOM software can be used to quickly generate random discrete observations. It was here used to create four orderings (n = 100) with units ranging from 1 to 7 (the rating scale). These four orderings were then averaged to create the final 25-unit (values of 1.0, 1.25, 1.50...6.50, 6.75, 7.00) outcome ordering.

Step 5. Run the analysis and save the randomization test results.

Step 6. Obtain the mean for the random-data PCCs. For this initial sample size of 100, the mean was equal to 46.58. A value of 50% might have been expected given pairs of observations are being compared to compute the PCC; however, ties are counted as incorrect classifications, thus lowering the PCC index.

Step 7. Examine the distribution of random-data PCCs. The full distribution of 1000 random-data PCCs from the randomization test was simplified and centered around the closeness range:

```
 0.00 :  41.57 : [186]****************
41.58:  51.58 : [636]****************************************************
51.59: 100.00 : [178]***************
```

As can be seen, ~64% of the random-data PCCs fell within the 41.58% to 51.58% range (±5% of the mean). A larger sample size is needed.

Step 8. Repeat the steps above until a suitable sample size is determined. The total sample size was doubled to 200, 100 per group. Running the analysis and saving the randomization results yielded a mean of 46.52 and the following simplified distribution:

```
 0.00:  41.51 : [120]**********
41.52:  51.52 : [771]*****************************************************************
51.53: 100.00 : [109]*********
```

As can be seen, ~77% of the random-data PCCs fell within the desired closeness range. The total sample size was again doubled to 400, 200 per group. Running the analysis and saving the randomization results yielded a mean of 46.70 and the following simplified distribution:

```
00.00 :  41.69 : [ 30]**
41.70 :  51.70 : [934]*****************************************************************
51.71 : 100.00 : [ 36]***
```

As can be seen ~93% of the values fell within the desired closeness range. Slightly larger sample sizes could be tried, but the 93% value is close enough to the desired 95% to conclude. Sawaoka and Monin would then proceed with their study, attempting to recruit 400 participants.

It is instructive to examine a more precise version of the random-data PCCs based on the

sample size of 400:

```
 0.00 :  3.99 : [  0]
 4.00 :  7.99 : [  0]
 8.00 : 11.99 : [  0]
12.00 : 15.99 : [  0]
16.00 : 19.99 : [  0]
20.00 : 23.99 : [  0]
24.00 : 27.99 : [  0]
28.00 : 31.99 : [  0]
32.00 : 35.99 : [  0]
36.00 : 39.99 : [  9]*
40.00 : 43.99 : [162]**************
44.00 : 47.99 : [508]****************************************
48.00 : 51.99 : [295]************************
52.00 : 55.99 : [ 26]**
56.00 : 59.99 : [  0]
60.00 : 63.99 : [  0]
64.00 : 67.99 : [  0]
68.00 : 71.99 : [  0]
72.00 : 75.99 : [  0]
76.00 : 79.99 : [  0]
80.00 : 83.99 : [  0]
84.00 : 87.99 : [  0]
88.00 : 91.99 : [  0]
92.00 : 95.99 : [  0]
96.00 : 100.0 : [  0]
```

As can be seen, no PCCs were observed in the top categories of the possible range of values. In fact, the maximum random-data PCC was equal to 55.29, a low value. As with traditional *p*-values, small and theoretically meaningless effects can be regarded as not plausibly due to chance when sample sizes are large. This fact goes a long way to explain why the observation oriented modeler treats the *c*-value as entirely secondary. The argument for plausibility rests primarily on the theoretically guided interpretation of the pattern in the data and the accompanying PCC index. For Sawaoka and Monin's actual study data, for instance, the PCC was equal to 54.59, with a *c*-value of .010. Without ties in the observations, a PCC equal to 50% would be expected under the assumption of physical chance. The *c*-value in this instance may help to rule out physical chance as an explanation of the pattern in the observations, but the low observed PCC value (just above 50%) unfortunately does not inspire confidence in Sawaoka and Monin's theory.

Finally, consistent with the results above based on randomly generated data, the randomization test conducted on Sawaoka and Monin's 390 participants yielded the following distribution of random-data PCCs:

```
 0.00 :  42.37 : [ 53]****
42.38 :  52.38 : [900]************************************************************
52.39 : 100.00 : [ 47]****
```

As can be seen, 90% of the PCCs fell within the ±5% range of the mean (47.38%). The estimated 93% for a sample of 400 participants was close to this observed percentage, indicating the OOM analog to the APP is sound.

Example #2

Suppose Sawaoka and Monin had designed their study with three groups in which participants read 2, 10, or 20 posted comments. Extending their prediction to these three groups, suppose they predicted the following ordinal relationship between the group ratings: $Grp_{20} > Grp_{10} > Grp_{2}$ (where "Grp" indicates "Group"). Given the average rating values ranging from 1 to 7, they could again use the *Ordinal Analysis – Crossed Orderings* procedure in the OOM software and obtain a PCC index for the complete, predicted pattern, $Grp_{20} > Grp_{10} > Grp_{2}$. Choosing the analysis is Step 1 in the process to determine sample size. Step 2 is the selection of the closeness value and desired closeness proportion. The same 5% closeness and 95% closeness proportion values as the first example were again chosen, and the starting sample size was set to 100 cases (50 persons per group). The *Generate Random Observations* procedure was again used in the OOM software to create the 100 observations. These choices account for Steps 3 through 5 above. Steps 6 through 8 were then followed until the sample size produced a distribution of random-data PCCs that met the closeness and closeness proportion values. Perhaps surprisingly, with the more complex ordinal pattern based on three groups rather than only two, the needed sample size was 180 participants (60 per group).

```
  6.52 :  8.56 : [ 18]**
  8.57 : 18.57 : [948]************************************************************************
 18.58 : 23.35 : [ 34]***
```

The mean for these random-data PCCs was equal to 13.57, and the lowest and highest values were 6.52 and 23.35, respectively.

Prior to collecting data for their study, Sawaoka and Monin could plan to examine the pairwise comparisons (viz., $Grp_{20} > Grp_{10}$, $Grp_{20} > Grp_{2}$, and $Grp_{10} > Grp_{2}$) between the three groups as well. If they were to do so, then they must choose the larger sample size of 400 participants from the first example above. Sampling only 180 participants will yield wider distributions of random-data PCCs for the pairwise group comparisons that will not match the expected 5% closeness and 95% closeness proportion values. For the complete 3group pattern, $Grp_{20} > Grp_{10} > Grp_{2}$, the larger sample size of 400 participants will yield more than 95% of the random-data PCCs in the 5% closeness range, which is a desirable effect on the distribution. In short, when confronted with varying desired sample sizes for different analyses, the largest sample size should be chosen.

Closing Comment

In closing, the steps presented above are to be performed with the understanding that this process is not an exact science, nor does it need to be an exact science. Tweaking the sample size in the steps above, for example, until the closeness percentage matches the desired target within .01% reveals a misunderstanding of how one should be thinking when using observation oriented modeling. Theory, patterns in the data, PCC indices, and exact replication all far exceed the *c*-value in importance. As stated above, the goal is to determine a sample size that will aid in your efforts to convince critics of the plausibility of your theory, and the steps above may prove helpful. The goal is not to estimate a population parameter with metric precision.