

REVIEW

User-friendly computer programs so econometricians can run the a priori procedure

Hui Li¹ David Trafimow^{2*} Tonghui Wang¹ Cong Wang¹ Liqun Hu¹

Abstract: In the 2019 special issue of *Econometrics* on significance testing and alternatives, Trafimow (2019) provided an alternative, termed the a priori procedure (APP). The APP involves finding the necessary sample size to meet prior specifications for precision and confidence and Trafimow reviewed equations for performing the APP. But the Trafimow article is limited in two important ways. Most important, the crucial equations must be solved by iteration, thereby rendering them impractical without the aid of relevant programming. The present work addresses the limitation by providing links to user-friendly programs, along with instructions, so even researchers unsophisticated in mathematics or statistics can use the APP. An additional limitation is that the APP bears a surface resemblance to power analysis. Although Trafimow had explained qualitatively why the APP and power analysis differ, there were no quantitative demonstrations. In contrast, the present article provides quantitative demonstrations to increase the clarity of the distinction. A conclusion that comes out of the quantitative demonstrations is that power analysis, as it is conventionally used, causes researchers to use insufficient sample sizes; an ironic conclusion as an important reason for researchers to perform power analyses is to address the problem of insufficient sample sizes. Thus, the present work is a follow-up piece to the previous *Econometrics* article because it addresses two important limitations of that article.

Keywords: a priori procedure, sample size, mean, difference in means, proportion, difference in proportions, program

1 Introduction

The invalidity of the null hypothesis significance testing procedure is being increasingly recognized. For example, *Basic and Applied Social Psychology* banned the procedure^[1,2]. More recently, the American Statistical Association produced a special issue of their flagship journal, *The American Statistician*, containing over 40 articles suggesting moving beyond significance testing because of its invalidity. Moreover, a strongly worded editorial, in that special issue, provided a clear repudiation of significance testing:

The ASA Statement on P-Values and Statistical Significance stopped just short of recommending that declarations of “statistically significance” be abandoned. We take

that step here. We conclude, based on our review of the articles in this special issue and the broader literature, that it is time to stop using the term “statistically significant” entirely. Nor should variants such as “significantly different”, “ $p < 0.05$ ”, and “non-significant” survive, whether expressed in words, by asterisks in a table, or in some other way^[3].

Several medical journals, such as the *New England Journal of Medicine* have changed their publication policies in accordance with the ASA statement.

Fields in economics or related to economics are changing too. Perhaps the most dramatic evidence of this is the *Econometrics* 2019 special issue on the topic. Trafimow (2019) published an article in the *Econometrics* special issue that not only advocates against significance testing; but advocates for a new inferential procedure, termed the a priori procedure (APP). Although the article provides relevant equations for applying the APP, an important limitation is the lack of accessible computer programming. Without such computer programming, researchers not expert in mathematics, statistics, and computer programming might find it difficult to use the APP, no matter its merits. The present article remedies the lack by providing links to programs that facilitate the ability of researchers to use the APP.

Received: March 19, 2020; **Accepted:** April 8, 2020; **Published:** April 9, 2020

Correspondence to: David Trafimow, Department of Psychology, New Mexico State University, Las Cruces, NM 88003-8001, USA; Email: dtrafimo@nmsu.edu

¹ Department of mathematical sciences, New Mexico State University, P.O. Box 30001, Las Cruces, NM 88003-8001, USA

² Department of Psychology, New Mexico State University, P.O. Box 30001, Las Cruces, NM 88003-8001, USA

Citation: Li H, Trafimow D, Wang T, *et al.* User-friendly computer programs so econometricians can run the a priori procedure. *Front Manage Bus*, 2020, 1(1): 2-6.

Copyright: © 2020 David Trafimow, *et al.* This is an open access article distributed under the terms of the [Creative Commons Attribution-Noncommercial 4.0 International License](https://creativecommons.org/licenses/by-nc/4.0/), which permits all non-commercial use, distribution, and reproduction in any medium, provided the original author and source are credited.

2 APP review and relevant websites

Unlike other inferential procedures, that are performed post-data; the APP is performed pre-data. The researcher considers two questions, pertaining to precision and confidence, before data collection^[4].

(1) Precision: How closely do I want the sample statistics of interest to approximate their corresponding population parameters?

(2) Confidence: With what probability do I want to attain the precision specification?

After committing to precision and confidence specifications, the researcher uses an appropriate APP equation to find the minimum sample size needed to meet those specifications. Once the required sample size, or a larger one, has been collected; no further inferential work is necessary. The researcher can compute the sample statistics of interest, confident that these sample statistics will be close to corresponding population parameters. There is no need for null hypotheses, p-values, or post-data confidence intervals.

For instance, consider the simplest case where a researcher assumes a normal distribution and collects a single sample of participants, with the goal of obtaining a sample mean that is within one-tenth of a standard deviation of the population mean, with a 95% probability. Equation (1) provides the simple solution:

$$n = \left(\frac{z_{(1-c)/2}}{f} \right)^2 \quad (1)$$

where n is the necessary sample size to meet specifications, f is the precision specification as a fraction of a standard deviation (0.1 in the example), and $z_{(1-c)/2}$ is the z -score that corresponds to the confidence specification (e.g., 1.96 for 95% confidence level in the example below).

Continuing with the example, the calculation would run as follows: $n = \left(\frac{1.96}{0.1} \right)^2 = 384.16$. Rounding up to the nearest whole number, as participants do not come in fractions, implies that the researcher needs to collect 385 participants to have a 95% chance of obtaining a sample mean within one-tenth of a standard deviation of the population mean.

Equation (1) is simple and does not really require a program. Nevertheless, we provide such a program at the website:

https://app-normal.shinyapps.io/N_SingleSample_EstimateMean_UnknownVariance

To use the program, type in the desired precision in the precision box and the desired level of confidence in the confidence box. Click the “Update” icon and the program will return the smallest sample size necessary to meet the

specifications.

Unlike Equation (1), where the standard deviation is assumed known, more complex APP equations feature inequalities, where the researcher has to find the smallest sample size for which the inequality is true. Let us consider a slightly modified example where the standard deviation is not known, in which case there is an inequality that involves the t -distribution with $n-1$ degrees of freedom, as opposed to the z -distribution we saw earlier:

$$t_{\alpha/2, n-1} \leq \sqrt{n}f \quad (2)$$

To obtain the required sample size without computer programming, the researcher would perform successive iterations to find the smallest value for n that satisfies the inequality. The researcher would try a value for n , check the result using the inequality, try another value for n , and eventually, after many iterations, converge on the smallest value for which the inequality is true.

Although, at the time of the Trafimow (2019) *Econometrics* article, no programs existed for carrying out Equation (2), this is no longer true. We now provide a practical alternative that the researcher can access the following website:

https://app-normal.shinyapps.io/N_SingleSample_EstimateMean_UnknownVariance

It is merely necessary is to type in the desired level of precision and the desired level of confidence and click the “Update” icon. The program will provide the sample size needed to meet specifications, provided that the precision varies between 0.20 and 0.90. For finer precision, the required sample size is extremely large, and so the t -distribution approximates the z -distribution. Thus, it is permissible to use the equation where the standard deviation is assumed known and the associated website provided earlier.

Sometimes researchers are interested in variances rather than means^[5]. For example, an econometrician might be interested in finding how much variance there is in wealth in the United States. To find the sample size needed to meet specifications for precision and confidence pertaining to variances, it is necessary to use the following equation:

$$\int_{L(n)}^{U(n)} f_{n-1}(u) du = c \quad (3)$$

where $f_{n-1}(u)$ is the density function of the chi-square distribution with degrees of freedom $n-1$.

For an easier alternative, use the program now available at the following website:

https://skewnormal.shinyapps.io/N_SingleSample_EstimateVariance/

Type in the desired precision and confidence levels, click the “Update” icon, and the program will return the necessary sample size to meet specifications.

The researcher may be interested in the difference in means for matched or independent samples^[6]. An example involving matched samples would be if an econometrician were interested in the increase or decrease in people’s income, in a particular population, over a decade. In a retrospective study, the researcher would select a sample of participants from the population and find their mean income a decade ago and today. The hope is that people’s mean incomes are better now than a decade ago. To determine the necessary sample size needed to meet precision and confidence specifications, it is necessary to use the following equation, which is the same as Equation (2); but with n referring to the sample size in both groups:

$$t_{\alpha/2, n-1} \leq \sqrt{n}f \quad (4)$$

An independent samples example would be if an econometrician were interested in comparing mean incomes in Illinois versus mean incomes in Indiana. There might be theoretical reasons to believe that incomes should be larger in one of the states than in the other, and the hope would be that the hypothesized difference would be supported by the data. When there are independent samples, there is no guarantee that the sample sizes will be equal, and it is convenient to designate that there are n participants in the smaller group and m participants in the larger group, where $k = n/m$. Using k , we derived Equation (5):

$$t_{\alpha/2, q} \leq \sqrt{\frac{n}{k+1}}f \quad (5)$$

where $t_{\alpha/2, q}$ is the critical t-score that corresponds to the level of confidence level $1-\alpha$ and degrees of freedom $q = n + \lceil \frac{n}{k} \rceil - 2$ in which $\lceil \frac{n}{k} \rceil$ is rounded to the nearest upper integer. Details on Equation (4) and Equation (5) are in Trafimow *et al.* (2020).

If the researcher has equal sample sizes, Equation (5) reduces to Equation (6):

$$t_{\alpha/2, 2(n-1)} \leq \sqrt{\frac{n}{2}}f \quad (6)$$

Alternatively, there is now an easier way. A researcher interested in the difference between means for either matched or independent samples, can access the following website:

https://app-normal.shinyapps.io/N_TwoSamples_EstimateMean

Simply type in the desired precision and confidence in the boxes provided and click the “Update” icon. The program will return the minimum sample size needed to

meet specifications for matched and independent samples. In the case of matched samples, of course, the two groups are the same size and contain the same participants. Thus, the sample size per group also equals the total sample size. But when samples are independent, the program returns the minimum sample size per condition. The sample sizes need not be the same in the two groups, but both groups must have the sample size returned by the program, or larger ones, to meet specifications. Another way to put this is that the smaller of the two groups must meet or exceed the sample size returned by the program.

Thus far, whether the statistic of concern was a mean, a variance, or a difference in means, we assumed normal distributions. However, APP equations also have been developed that pertain to proportions or differences in proportions. For example, a researcher might wish to estimate the proportion of people that can afford a new car. Or, a researcher might wish to compare the proportion of people that can afford a new car in New Mexico with the proportion of people that can afford a new car in Arizona. Either way, the distribution is binomial, rather than normal, with the important caveat that with a sufficiently large number of participants, the two distributions approximate each other. In the case where there is a single sample and the researcher wishes to know how many participants to collect to meet specifications for precision and confidence, the following equation can be used:

$$n = \frac{z_{\alpha/2}^2 p_0 q_0}{E^2} \quad (7)$$

An easier way is now available at the following website: https://app-normal.shinyapps.io/EstimateProp_OneSample

To use the website, type in the precision and confidence specifications, as before. But there are additional issues. First, unlike the previous cases, precision is not specified as a fraction of a standard deviation; but rather as an absolute number. For example, specifying 0.03 for precision implies that the sample proportion will be within 0.03 of the population proportion at the specified level of confidence. Second, there is an additional box, which is an initial estimate p_0 of the proportion (note that $q_0 = 1-p_0$) from previous data information. If there is a good reason for assuming a value other than 0.50, such as the availability of previous data; type in that value. Otherwise, use 0.50. After clicking “Update”, the program will return the necessary sample size.

Or, if the researcher is interested in comparing two proportions, the difference in sample proportions is used to estimate the difference in population proportions. To find the minimum sample size needed to meet specifications for precision and confidence, the researcher can use the

following equation:

$$n \leq \frac{z_{\alpha/2}^2 (p_{10}q_{10} + p_{20}q_{20})}{E^2} \quad (8)$$

We provide a program at the follow website:

https://app-normal.shinyapps.io/EstimateProp_TwoSamples

To use the program, specify precision and confidence, as described earlier with respect to a single proportion. In addition, there are two extra boxes that pertain to the two samples, respectively, where p_{10} and p_{20} are initial estimates of the proportions from previous information of two independent samples. If there is a good reason to have a particular estimate, with respect to one or both of the samples, type in the estimates in the boxes provided. If not, use 0.50 in one or both boxes.

3 How the APP and power analysis differ

Because the goal of the APP is to help researchers determine the sample sizes to use and the goal of power analysis also is to help researchers determine the sample sizes to use, there is a surface similarity between the APP and power analysis. However, the two are extremely different and one way to demonstrate the usefulness of the computer programs is to use them to for quantitative contrasts.

Let us begin qualitatively. To reiterate, the goal of the APP is to find the sample sizes researchers need to meet specifications for precision and confidence. In contrast, the goal of power analysis is to find the samples sizes researchers need to have a good chance of obtaining statistical significance at a specified effect size. Typically, though not always, researchers assume a medium size effect (*e.g.*, Cohen's $d = 0.05$) and 80% power. The following demonstrations will make use of these conventions. The important qualitative point is that the APP is highly influenced by the desired level of precision, whereas power analysis is not; but power analysis is highly influenced by the designated effect size, whereas the APP is not.

Let us now use the computer programs for some quantitative contrasts. First, imagine the case of a single group, where the researcher wishes to obtain a sample mean that is a good estimate of the population mean, such as obtaining a sample of Americans with the goal of obtaining a sample mean to estimate the population mean. Remaining with conventions, the researcher wishes to have 80% power to determine a medium size effect; and uses a power analysis calculator to find that the necessary sample size is 31. However, using the single sample APP program with known variance implies that the precision is an unimpressive 0.35 at the conventional 95% confidence level. The sample mean to be obtained has a 95% prob-

ability of being within 0.35 standard deviations of the population mean, in either direction, for a total interval of 0.70. Assuming unknown variance leads to a slightly worse precision value of 0.36. Either way, the unimpressive precision value demonstrates that the conventional use of power analysis leads to imprecise research.

Nor must the demonstration be limited to the simple case of a single sample. Let us consider the slightly more complex case where the econometrician is interested in a difference between means (matched samples), such as the difference in income now relative to a decade ago. How many participants should the econometrician collect? According to power analysis, the researcher needs to collect 33 participants to detect a medium effect size with 80% power. But using the APP program pertaining to differences in means indicates that the precision level is only 0.35, so that the total precision interval is 0.70. Nor are matters improved if we consider the difference in means in independent samples, such as the difference in mean income between residents of Illinois and Indiana. According to conventional power analysis, the researcher would need to collect 63 participants from each state to detect a medium effect size with 80% power. However, the precision level would be 0.35, for a total precision interval of 0.70. Again, we see that conventional power analysis leads researchers to conduct studies with poor precision.

Nor do matters change when we consider a difference in proportions. Suppose the researcher wished to detect a medium size effect at 80% power (for proportions, the effect size statistic would be h rather than d). For example, a difference in proportions between 0.68 and 0.43 would result in $h \approx 0.5$. According to power analysis, the necessary sample size would be 63 participants in each condition. But using the APP program concerned with differences in proportions, this leads to precision equals 0.17 and a total precision interval of 0.34. As we explained earlier, unlike the examples involving differences in means, when differences in proportions are involved, the precision interval is an absolute number rather than a fraction of a standard deviation. We leave it to the reader to judge whether power analysis performs worse, in precision terms, in the context of differences in means versus differences in proportions.

4 Discussion

To render the APP easier to use, we provided free and publicly available computer programs, along with descriptions of how to use them. For the reader's convenience, all links can be accessed in [Table 1](#). The user merely pastes the appropriate link into her browser to use the chosen program.

Table 1. Publicly available computer programs that can be used to find the required sample size for different researcher goals. Note that the link for 'difference in means' works for both matched and independent samples

Researcher Goal	Website
Mean for single sample, known variance	https://app-normal.shinyapps.io/N_SingleSample_EstimateMean_KnownVariance
Mean for single sample, unknown variance	https://app-normal.shinyapps.io/N_SingleSample_EstimateMean_UnknownVariance
Variance of single sample	https://skewnormal.shinyapps.io/N_SingleSample_EstimateVariance
Difference in means	https://app-normal.shinyapps.io/N_TwoSamples_EstimateMean
Proportion	https://app-normal.shinyapps.io/EstimateProp_OneSample
Difference in proportions	https://app-normal.shinyapps.io/EstimateProp_TwoSamples

As always, there are limitations. The main limitation here is the lack of consideration of skew-normal distributions. The family of normal distributions is a subset of the family of skew-normal distributions. Therefore, the family of skew-normal distributions is generally more applicable than the family of normal distributions, though the family of skew-normal distributions is much more difficult to handle. Consequently, there are APP equations pertaining to skew-normal distributions in various stages of progress, analogous to the present equations pertaining to normal distributions. To address the limitation, we expect to write a companion piece, similar to the present one, but that can handle skew-normal distributions.

An interesting point that Trafimow, Wang and Wang (2019)^[7] demonstrated is that skewnormal distributions imply more precision than do normal distributions. A consequence is that the present APP equations pertaining to normal distributions are more conservative than APP equations pertaining to skew-normal distributions, in the sense of requiring larger sample sizes to meet specifications for precision and confidence. Thus, if the researcher does not know the skewness of the distribution or distributions of interest, the assumption of normality can be considered a conservative way to address the lack of knowledge. That is, the sample sizes produced by the present programs will meet or exceed the necessary sample sizes to meet specifications for precision and confidence when there is skewness. Therefore, for researchers who wish to act conservatively, the present programs should suffice even for skewed distributions.

We hope and expect that the present programs will facilitate econometricians using the APP. No particular mathematical or statistical ability is required to use the programs. Nor is there a requirement for the user to be expert at programming, to know any programming languages, or even to know any statistical packages. However, to use the programs appropriately, it is of utmost

importance to understand the APP philosophy. Towards this end, we recommend that users consult the companion *Econometrics* article by Trafimow (2019)^[8], that spells out that philosophy in greater detail than the present article.

References

- [1] Paoletta MS. Linear Models and Time-Series Analysis: Regression, ANOVA, ARMA and GARCH, Chichester: John Wiley and Sons, 2018.
- [2] Trafimow D. Using the Coefficient of Confidence to Make the Philosophical Switch From A Posteriori to A Priori Inferential Statistics. *Educational & Psychological Measurement*, 2017, **77**: 831-854. <https://doi.org/10.1177/0013164416667977>
- [3] Trafimow D. A frequentist alternative to significance testing, p-values, and confidence intervals. *Econometrics*, 2019, **7**(2): 1-14.
- [4] Trafimow D and Michael M. Editorial. *Basic and Applied Social Psychology*, 2015, **37**: 1-2.
- [5] Trafimow D, Wang T and Wang C. From a sampling precision perspective, skewness is a friend and not an enemy! *Educational and Psychological Measurement*, 2019, **79**(1): 129-150. <https://doi.org/10.1177/0013164418764801>
- [6] Trafimow D, Wang T and Wang C. Making the a priori procedure (APP) work for differences between means. *Educational and Psychological Measurement*, 2020, **80**: 186-198. <https://doi.org/10.1177/0013164419847509>
- [7] Wang C, Wang T, Trafimow D, *et al.* Necessary sample size for estimating the scale parameter with specified closeness and confidence. *International Journal of Intelligent Technologies and Applied Statistics*, 2019, **12**: 17-29.
- [8] Wasserstein RL, Schirm AL and Lazar NA. Moving to a world beyond "p < 0.05". *The American Statistician*, 2019, **73**(51): 1-19. <https://doi.org/10.1080/00031305.2019.1583913>