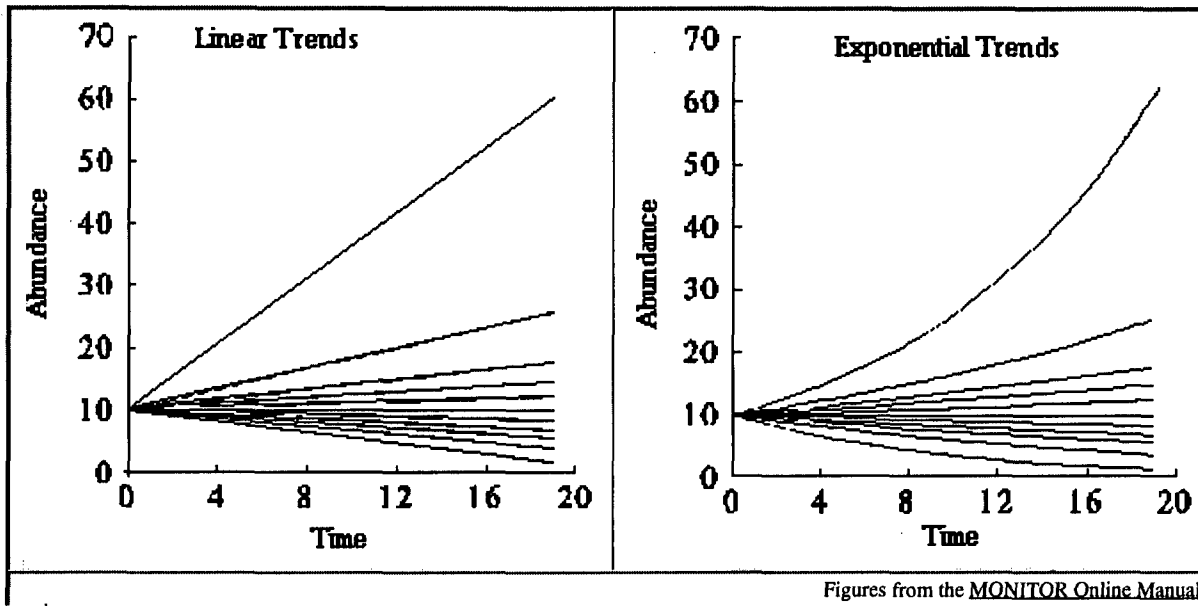


Trend Type & Trend Level



Trend Type



What type of trend underlies changes in your counts? Two basic models are linear vs. exponential growth. If your counts are more or less normally distributed and if trends in your counts are expected to be linear in nature, choose the linear model. Choose the exponential model if your counts are log-normally distributed and if trends in the population monitored are likely exponential in nature. It is generally thought that the log-normal distribution best approximates the growth of most animal populations.

Trend Level

What magnitude of population change would you like your program to be able to detect? This question is often answered in terms of percent change over a given time period. For example, the Terrestrial Salamander Monitoring Program sampling frame calls for the detection of at least a 50% decline over 20 years, which translates into a -3.4% trend *per year*. All things being equal, you will need more samples to detect smaller trends over shorter time intervals. For comparative purposes it is usually better to express trend on a per year basis rather than as total trend. The following table will help in converting between long- and short-term trends.

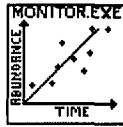
Short-term vs. long-term trends. For those of us who steered clear of finance class, percent change over time can be a tricky concept. Here's a chart that I whipped together using the Converter sub-program in *MONITOR*. As you can see, a 5% decline per year is a deceptively critical drop as the years stack up!

Trend per year →	-5%	-3%	-2%
Trend over 5 years	-22.62%	-14.13%	-9.61%
Trend over 10 years	-40.13%	-26.26%	-18.29%
Trend over 15 years	-53.67%	-36.67%	-26.14%
Trend over 20 years	-64.15%	-45.62%	-33.24%



Back to Power Analysis Step by Step

Page developed by Paige C. Eagle
 For more information please contact:
 Sam Droege (frog@usgs.gov)
 USGS Patuxent Wildlife Research Center
 12100 Beech Forest Road
 Laurel, MD 20708-4038
 301-497-5840



Selecting Precision Criteria for a Monitoring Program



Level of Significance (α). Think of the significance or alpha level as the percentage of time you are willing to "cry wolf" and incorrectly conclude that a population has undergone significant change when it actually has not. There is a direct trade-off between significance level and the power of a study to detect population changes that are occurring. A statistical test that is significant at the 5% level of significance ($\alpha = 0.05$) may not be significant at the 1% level. As alpha decreases, there is a corresponding decrease in power .

We here at NAAMP feel that an **alpha level of 0.10–0.20** is well-suited to the goals of monitoring programs. In most experimental settings when we make a statement that there is a "statistically significant difference" between or among things that we have measured, we want to be positive those differences are real, otherwise our credibility suffers. However, when monitoring populations (as opposed to running experiments on populations), the situation is different. The consequences of sounding a false alarm are small when compared to the consequences of failing to detect a population crash because we set our standards too high (by setting our alpha level too low). As conservationists, our job is to catch these events when extinction/extirpation might be prevented. Consequently, the penalty of possibly being wrong 10–20% of the time in our statements regarding the presence of "significant change" is well worth the aggravation of sometimes crying wolf. The alternative choice is to let a population slip between the cracks (cracks caused by variability in our counting techniques) because we waited too long for the numbers to tell the story.

Power ($1 - \beta$). In terms of a monitoring program, power is the ability to detect a trend if one is really going on. For the same reasons detailed above, **we believe a monitoring program should be powerful enough to detect a real population trend 90% of the time it actually is occurring.** This is what monitoring is all about, so let's play to win.



[Back to Power Analysis Step by Step](#)

Page developed by [Paige C. Eagle](#).
For more information please contact:
[Sam Droege \(frog@usgs.gov\)](mailto:frog@usgs.gov)
[USGS Patuxent Wildlife Research Center](#)
12100 Beech Forest Road
Laurel, MD 20708-4038
301-497-5840



Power Step Analysis: by Step

Power Analysis Step-by-Step

To estimate the sampling effort needed to ensure our study is *powerful* enough to detect the trends we are interested in (should they actually be occurring) requires the following steps:

1. Decide what **type and level of trend** you want your program to detect.
2. Decide **how precisely** you want to talk about these trends.
3. Figure out **how variable** the technique is for the species you want to monitor. (Need variance in counts over time on the same plot--*not* spatial variance in counts among plots). For some published values, see the [Variability Estimates for Local Populations of Plants and Animals](#) page.
4. Use the values from steps 1-3 to determine the **level of sampling effort required**:
 - A. By referring to our site [How many, how often, and how long?](#)
 - B. By running the freeware program [MONITOR](#) tailored to your needs.
5. Finally, while not required for the determination of sample sizes, you will need to consider the **bias** inherent in the sampling technique (or its application). How accurately does the estimated trend track the true population trend? A program can have counts with low variance and excellent power, but have little relationship with the actual population status of the organism being counted. Validation is the procedure by which an abundance index is compared to true estimates or known values of population size.

See also: [Terrestrial Salamander Monitoring Program Bias Page](#), [A Heuristic Approach to Validating Monitoring Programs Based on Count Indices](#), and our [Validation of wildlife abundance indices bibliography](#).



(Last updated: December 20, 1999)

Page developed by [Paige C. Eagle](#).

For more information please contact:

Sam Droege (frog@usgs.gov)

[Patuxent Wildlife Research Center](#)

12100 Beech Forest Road, Laurel, MD 20708-4038

301-497-5840



How... *many, often, & long?*

Optimizing sampling effort for successful monitoring programs

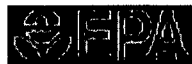
How many, how often, & how long?

Here's where we try to answer those burning questions we all must face when planning any study: how many samples will I need? In a monitoring program, sample size technically refers to survey length, but other types of sampling effort include number of study plots (routes, subpopulations) and number of counts per survey period (year, usually). Increasing these latter two variables increases replicates over space and time, thus increasing precision and power and decreasing required sample size.

First, what's your question?

How many? I am more limited by how many replicate counts I can make each year than by number of sites I can monitor. (E.g., you've got plenty of potential study plots, but funds are such that you can hire a bunch of folks for only a week each year, or you're using volunteers who can't make the time commitment that repeated counts requires). You need to know *how many* plots to survey.

How often? I am more limited by number of study sites than by counts per year. (E.g., could be that there's only one pond on your refuge, or travel between sites is prohibitively expensive). You need to know *how often* to count per year.



(Last updated: December 20, 1999)

Page developed by Paige C. Eagle.

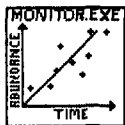
For more information please contact:

Sam Droege (frog@usgs.gov)

Patuxent Wildlife Research Center

12100 Beech Forest Road, Laurel, MD 20708-4038

301-497-5840



Sampling Frame for Running Power Simulations using MONITOR to develop graphs for "How Many?"



The following sampling frame describes ranges of inputs used in the program MONITOR to develop graphs for our page "How Many?" To better understand the terminology as it applies to MONITOR and relates to a real world study, we *highly* recommend perusing the [MONITOR Online Manual](#) before using the graphs.

Constants:	Variables:
<ul style="list-style-type: none"> • Power <math>\geq 90\%¹</math> • Alpha level = <math>0.20²</math> • Variability in trend among sites: standard deviation = mean trend (0.05, 0.03, or 0.02 depending)* • Two-tailed test • Exponential growth • Whole number rounding • 1000 iterations 	<ul style="list-style-type: none"> • Survey lengths: 5, 6, 7, 8, 9, 10, 15, 20, 50 years • Counts/plot/survey: 1, 3, 5 • Coefficients of Variation (CV's): 20%, 40%, 60%, 80%, 100%, 125%, and 150% • Levels of trend detection: -5%, -3%, -2% per year* <p>* Corresponds to 64%, 46%, and 33% declines over 20 years, respectively.</p>
<p>*A conservative estimate of trend variation corresponding to an entry of 1 for the value of trend CV.</p>	<p>More information about short-term vs. long-term trends.</p>



This page is a work in progress. **We are still fine-tuning the methods surrounding power analysis and estimation of variance in counts over time. Here we provide our most current sample size estimates, which are based on many simplifying assumptions needed to run simulations. In short, the suggestions you'll find here should be viewed not as absolutes, but as approximations from which to start.**

These are minimum statistical precision goals for each site. The expectation is that most sites would far exceed these targets. There are many reasons why these sample sizes should be viewed not as absolutes, but as the barest of bare minimums from which to start:


- Since *MONITOR* generates power estimates from stochastically driven simulations, the output changes from run to run.
- These simulations were run under an alpha level of $0.20²$. If your study design calls for a more stringent level of significance, *you will require more samples.*
- These results assume a coefficient of variation in trend of 1.0 among sites surveyed. Between-route variation has an inverse effect on power and if you think population trends vary more or less among your sites, you'll need to adjust your sample size.

In short, it would be prudent to pad the numbers upwards whenever possible, and reevaluate your program using *MONITOR* once you have several years of data.

¹**Power = 90%. Power is the ability of a monitoring program to detect a trend if one is really going on. We believe a monitoring program should be powerful enough to detect a real population trend 90% of the time it actually is occurring.**

²**Alpha level = 0.20. Setting the alpha to this level means that we are willing to cry wolf approximately 20% of the time. In other words, we are willing to incorrectly identify a population as one having undergone a significant change, when in reality it had not, 20% of the time.**

Why 20% rather than the traditional 5%? The reason lies in our goals. In most experimental settings when we make a statement that there is a "statistically significant difference" between or among things that we have measured, we want to be damn sure that those differences are real, for if they are not, we look foolish, and people lose respect for us and our interpretation of biology. However, when monitoring populations (as opposed to running experiments on populations), the situation is different. It is imperative that we detect any important changes in populations, because as conservationists we sure would look bad if the population crashed over the span of a few years, and we missed the start of the event (when extinction/extirpation might possibly have been prevented) because we set our standards for detecting trends too high (by setting our alpha level too low). Consequently, the penalty of possibly being wrong 20% of the time in our statements regarding the presence of "significant change" is well worth the aggravation of sometimes crying wolf. The alternative choice is to let a population slip between the cracks (cracks caused by variability in our counting techniques) because we waited too long for the numbers to tell the story.

 **[Back to How many plots?](#)**

Page developed by [Paige C. Eagle](#).

For more information please contact:

[Sam Droege \(frog@usgs.gov\)](mailto:frog@usgs.gov)

[USGS Patuxent Wildlife Research Center](#)

12100 Beech Forest Road

Laurel, MD 20708-4038

301-497-5840

Getting the Sample Size Right:

A Brief Introduction to Power Analysis

[Jeremy Miles Home Page](#)
[Miscellany Page](#)

Queries: [Email Jeremy Miles](#)

Introducing Power Analysis

A hypothesis test tells us the probability of our result occurring, if the null is false. If the probability is lower than a pre-specified value (alpha, usually 0.05), the null hypothesis is rejected. It can be likened to a search process, we are searching for evidence to reject the null hypothesis, in the same way that we may search for (say) a presen chemical in an environment.

The ability to reject the null hypothesis depends upon:

- **Alpha (α):** Usually set to be 0.05, although this is somewhat arbitrary probability of a type I error, that is the probability of rejecting the null hypothesis given that the null hypothesis is true. To use the search analogy, it is the probability of thinking we have found something when it is not really there.
- **Sample size:** A larger sample size leads to more accurate parameter estimates, which leads to a greater ability to find what we were looking for. The more we look, the more likely we are to find it.
- **Effect Size:** The size of the effect in the population. The bigger it is, the more likely it will be to find.

However, the above is not strictly correct. Jacob Cohen (author of several articles on power analysis) has pointed out *"all null hypotheses, at least in 2-tailed forms, are false."* Whatever we are looking for is always going to be there in such small quantities that we are not bothered about it.

Power analysis allows us to make sure that we have looked hard enough to find it, if there is enough of it there to bother us. The thing we are looking for is known as the *"effect size."* Several methods exist for deciding what effect size we would be interested in. Different statistical tests have different effect sizes developed for them, however the general principle is:

- **Base it on substantive knowledge.** Kraemer and Thiemann (see end) give the following example. It is hypothesised that 40 year old men who drink

three cups of coffee per day will score more highly on the Cornell Men (CMI) than men who do not drink coffee. The CMI ranges from 0 to 1 previous research has shown that scores on the CMI increase by about for every decade of life. It is decided that an increase, caused by drink which was equivalent to about 10 years of age would be enough to w concern, and so an effect size can be calculated based on that assum

- Base it on previous research. See what effect sizes other researchers similar fields have found. Use this as an estimate of sample size. Use conventions. Cohen (again) has defined small, medium and large for many types of test. These form useful conventions, and can guide know approximately how strong the effect is likely to be.

Doing Power Analysis

Three types of power analysis exist, *a priori*, *post hoc*, and compromise. 1 workshop will focus on the first two. Compromise power analysis is a mo issue, which is rarely used, and slightly controversial.

1. A Priori Power Analysis

Ideally, power analysis is carried out *a priori*, that is during the design sta study. A study can conclude that a null hypothesis was true or false. The i.e. the state of the world, can be that the hypothesis is true or false.

Given the three factors alpha, sample size and effect size, a fourth variabl calculated, called beta. Where alpha is the probability of a type I error (i.e of a correct null hypothesis) beta is the probability of a type II error (acc false null hypothesis).

		State of the World	
		H_0 True	H_0 Fals
Research Findings	H_0 True	✓	Type II Er ($p = \beta$)
	H_0 False	Type I Error ($p = \alpha$)	✓

The probability of correctly accepting the null hypothesis is equal to $1 - \alpha$, fixed, the probability of incorrectly rejecting the null hypothesis is β . The probability of correctly rejecting the null hypothesis is equal to $1 - \beta$, which is called power of a test refers to its ability to detect what it is looking for. To return to the search analogy, the power of a test is our probability of finding what we are looking for, given its size.

A power analysis program can be used to determine power given the value of sample size and effect size. If the power is deemed to be insufficient, steps can be taken to increase the power, (most commonly, but not exclusively, by increasing sample size.)

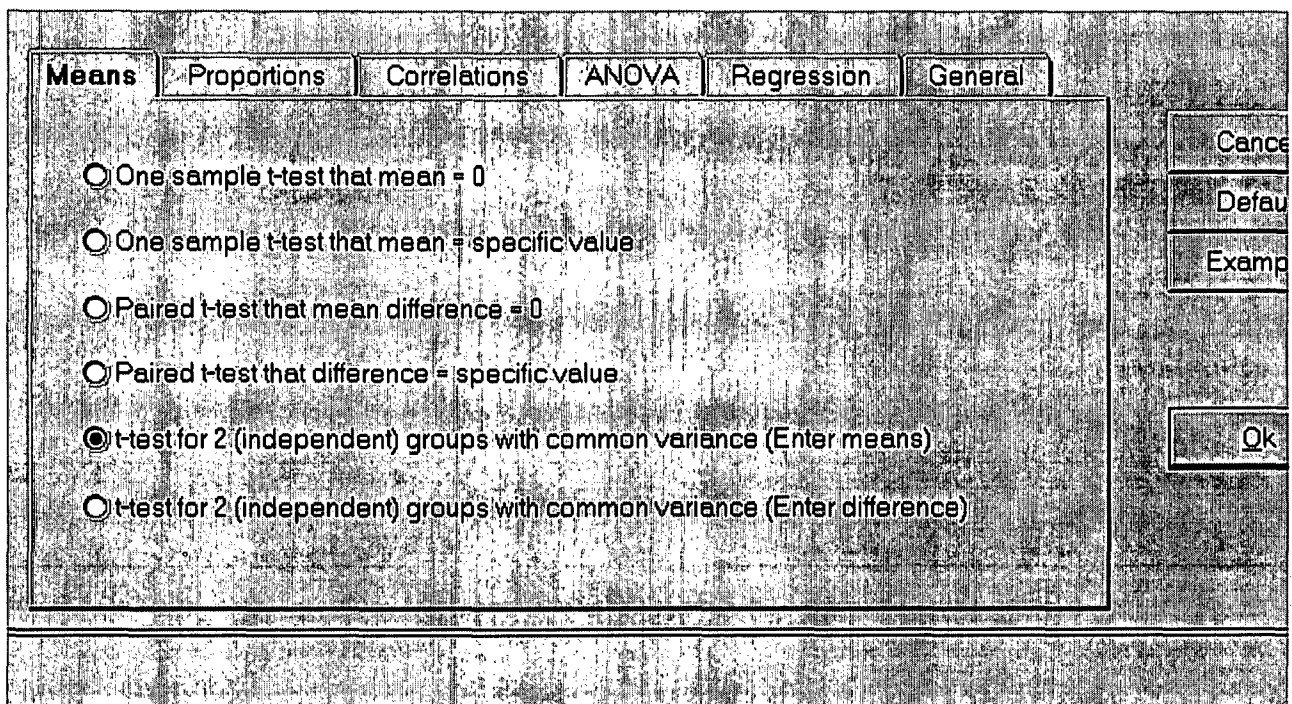


Figure 1: Picking an Appropriate Test

Example Using SamplePower

We will use the previous example of Kraemer and Thiemann, in which we were interested in an increase, approximately equivalent to 1 decade, caused by coffee. Figure 1 shows the first screen of the SamplePower program in which we select the program what sort of statistical test we will be using.

In this case it is a t-test with 2 independent groups. We know that the average CMI for men in their 40's is 8, and that the SD is 7. We have decided that an increase of 3.5 (increase associated with ageing 1 decade) will be sufficient.

concern us. (This increase would, in Cohen's classifications, be a medium

	Population Mean	Standard Deviation	N Per Group	Standard Error	95% Lower
Population 1	8.0	7.0	51		
Population 2	11.5	7.0	51		
Mean Difference	-3.5	7.0	102	1.36	1.21
Alpha=0.05 Tails= 1					Power

Figure 2: Sample Size Calculation for 80% Power

Figure 2 shows the sample size for 80% power (usually considered) to be power. This would require an N of 51 or group, or a total N of 102.

A graph can be a useful way of making the best decision regarding the trade between power and sample size. The graph drawn by SamplePower is shown in Figure 3. It can be seen that the rate of increase of increase in power starts to reduce dramatically at around the 80% power figure. An appropriate sample size can be determined, based on expense and desired power.

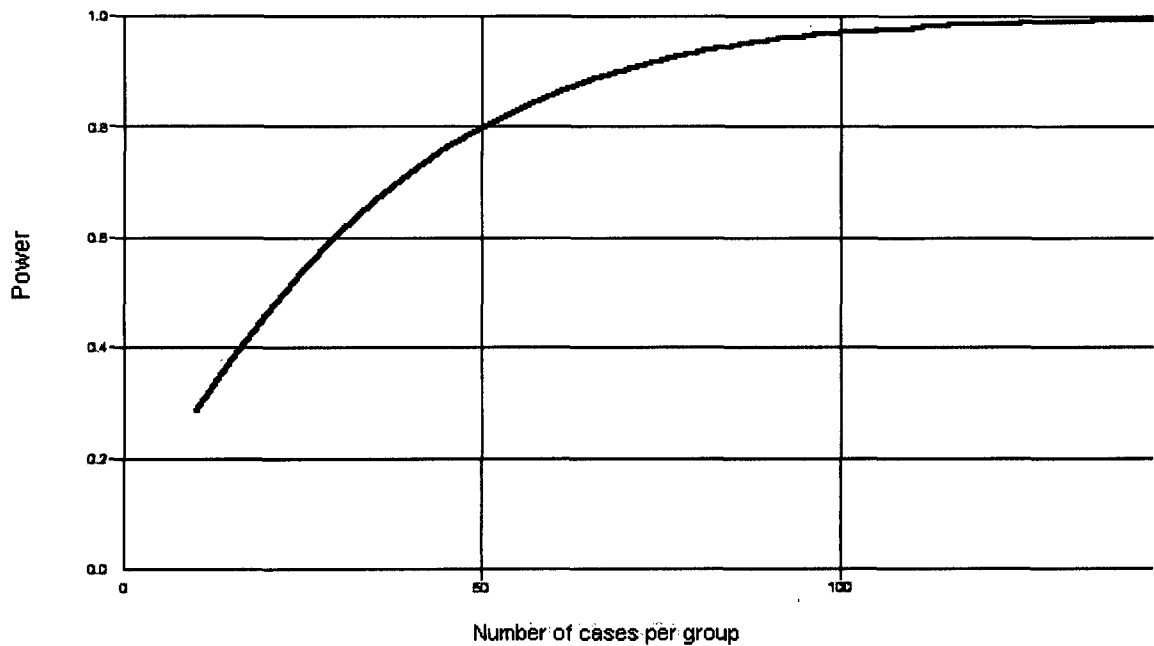


Figure 3: Graph Drawn by SamplePower

A final and very useful feature of SamplePower is its ability to generate a report, to communicate the findings of the power analysis.

Power for a test of the null hypothesis

One goal of the proposed study is to test the null hypothesis that the two population means are equal. The criterion for significance (alpha) has been set at 0.05. The test is 1-tailed, which means that the expected direction will be interpreted.

With the proposed sample size of 51 and 51 for the two groups, the study will have power to detect a statistically significant result.

This computation assumes that the mean difference is -3.5 (corresponding to means of 8 and 11.5) and the common within-group standard deviation is 7.0.

This effect was selected as the smallest effect that would be important to detect, in the sense that a smaller effect would not be of clinical or substantive significance. It is also assumed that the effect is reasonable, in the sense that an effect of this magnitude could be anticipated in this field.

Precision for estimating the effect size

A second goal of this study is to estimate the mean difference between the two populations. A study of this design would enable us to report the mean difference with a precision (95.0% confidence interval) of plus/minus 2.29 points.

For example, an observed difference of -3.5 would be reported with a 95.0% confidence interval from minus infinity to 5.79, or (alternatively, per the a priori hypothesis) of minus infinity to 5.79. (Since the confidence interval has been defined as one tailed, only one boundary is meaningful).

The precision estimated here is the median precision. Precision will vary as a function of standard deviation (as well as sample size), and in any single study will be narrower or wider than the estimate.

Notes

Computational option: Variance is estimated (t-test)

A priori power analysis can ensure that you do not waste time and resources carrying out a study which has very little chance of finding a significant effect and can also ensure that you do not waste time and resources testing more subjects than are necessary to detect an effect.

Post Hoc Analysis

Whereas a priori analysis is done before a study has been carried out, a post hoc analysis is done after a study has been carried out to help to explain the results of a study which did not find any significant effects.

Imagine that a study had been carried out to see if CMI scores were correlated with coffee consumption (measured in average number of cups per day.) A researcher carries out a study, using 40 subjects, fails to find a significant correlation, and therefore concludes that coffee consumption does not affect CMI score.

The effect size for correlation coefficients is simply r , the correlation coefficient. A power analysis can be carried out to find out what effect size it would be likely to detect. This is most easily done by examining a graph of power as a function of effect size, for the sample size used in the study. This graph is shown in Figure 4. The x-axis goes from 0.1 to 0.5. (Cohen defines a small effect size to be $r=0.1$, a medium effect size to be $r = 0.3$, and a large effect size to be $r = 0.5$.)

It can be seen from the graph that the power to detect a large effect size is high, above 0.95. It could therefore be safely concluded from this study that a large effect size is present.

coffee consumption does not have a large effect on CMI score. At the effect size ($r = 0.3$) the power to detect a significant result is slightly Although this is reasonable power, it is not sufficient to safely assume is not a medium effect. Finally at a small effect size, the power is very around 0.15. This is certainly not sufficient to conclude that there is r effect.

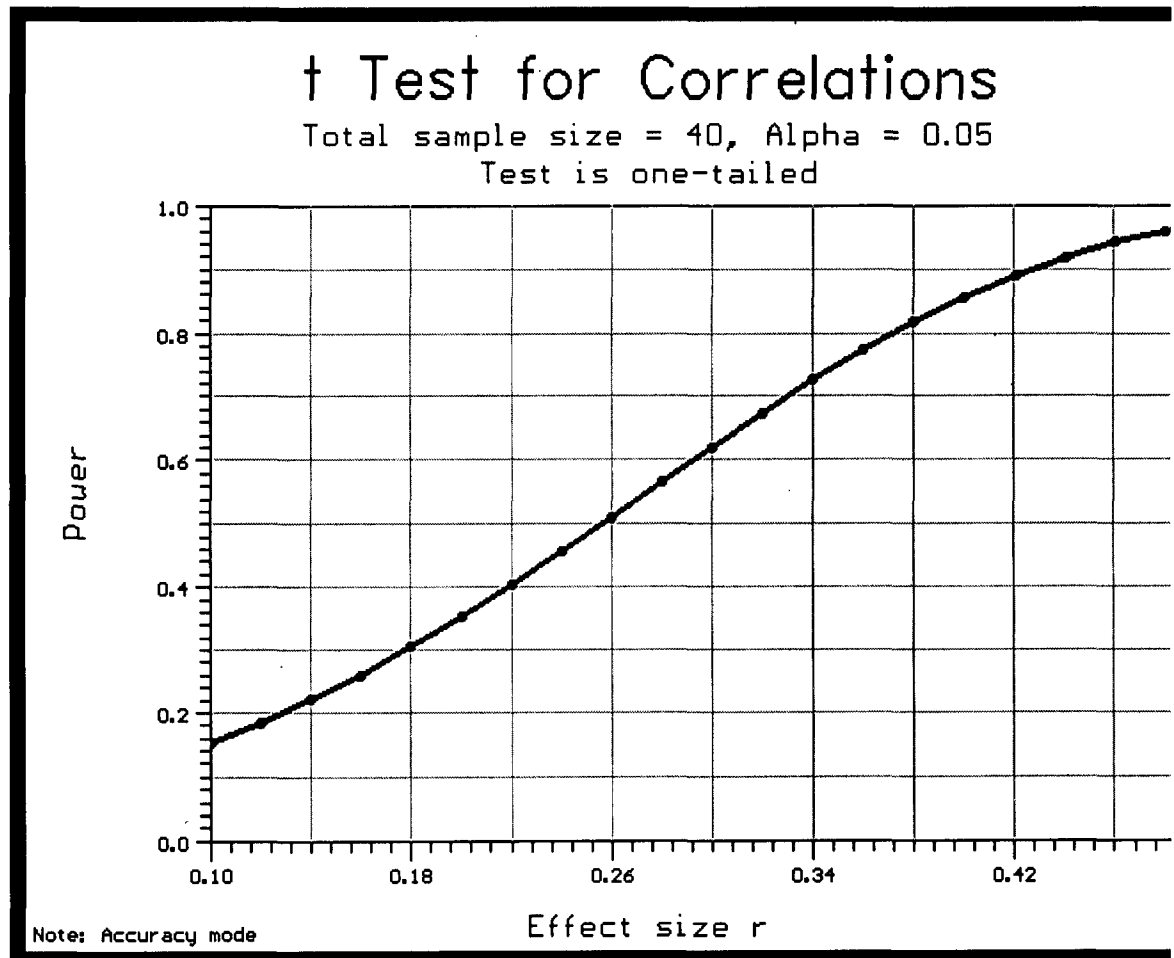


Figure 4: Power as a function of effect size

Whether the power of the study was sufficient to decide that there is

the population would depend upon the degree of correlation which was determined to be large enough to be important.

Increasing Power

Although studies with excessive power exist, they tend to be few and between. The main problem in designing and carrying out studies is to have sufficient power.

1. Increase Sample Size

As we have seen, the main way of increasing power is to increase sample size.

2. Increase Alpha

All of the studies we have mentioned have used an alpha of 0.05. An alpha of 0.05 disregards the role of alpha in determining the level of significance.

If we assume a correlation value of 0.3 (medium effect size) to be the true finding:

$\alpha = 0.05$ (2-tailed), $N = 40$, Power = 0.49

$\alpha = 0.10$ (2-tailed), $N = 40$, Power = 0.62

3. Shrink Standard Deviations

By using more homogenous groups (in an experimental study) the relative error size increases. Similarly increasing the reliability of the measures will have the same effect.

4. Use ANCOVA

Adding covariates to an experimental study statistically reduces the error variance, and therefore increases the relative effect size.

In the following table SD is always equal to 1, $\alpha = 0.05$

Mean Group 1	Mean Group 2	Effect Size	Total N	R ² of Covs
0.0	0.2	Small	100	0.00
0.0	0.2	Small	100	0.49
0.0	0.2	Small	200	0.00
0.0	0.5	Medium	60	0.00
0.0	0.5	Medium	60	0.25
0.0	0.5	Medium	80	0.00

Addition of covariates with an R² of 0.49 (i.e. a correlation of 0.7 with increases power to the same extent as a doubling in sample size. Add covariates with an R² of 0.25 (correlation of 0.5) increases power to extent as an increase in sample size of one third.

Appendix 1: Further Reading

Books about Power Analysis

Cohen, J. (1992). *Statistical Power Analysis for the Behavioural Sciences*. Hillsdale, NJ: Erlbaum. ISBN: 0-8058-0283-5

Kraemer, H.C. and Theimann, S. (1987). *How many subjects? Statistical analysis in research*. Newbury Park, CA: Sage. ISBN: 0-8039-2949-8

Books Which Discuss Power Analysis (You can see pricing information at An by clicking on the title.)

Abelson, R.P. (1995). *Statistics as principled argument*. Hillsdale, NJ: ISBN: 0-8058-0528-1

Chow, S.L. (1996). *Statistical Significance: Rationale, Validity and Ut* Sage. ISBN 0-7619-5205-5

Several chapters in the two volume set:

Keren, G. and Lewis, C. (1993). *A Handbook for Data Analysis in the*

Sciences. Hillsdale, NJ: Erlbaum.

Or separately:

Methodological Issues ISBN: 0-8058-1037-4

Statistical Issues ISBN: 0-8058-1093-5

These include:

Gigerenzer, G. *The superego, the ego and the id in statistical reasoning*. (Methodological Issues)

Tversky, A. and Kahneman, D. *Belief in the law of small numbers*. (Methodological Issues – this is an edited reprint of their classic article from the *Psychological Bulletin*).

Greenwald, A.G. *Consequences of prejudice against the null hypothesis*. (Methodological Issues).

Tatsuoka, M. *Effect Size*. (Methodological Issues).

Appendix 2: Computer Programs

Power analysis calculations are fairly difficult to carry out, they often have no analytical solution. It is possible to calculate power using the non-central distribution functions in SPSS, although not easy. The two computer programs used to generate the results for this paper were GPower and SamplePower.

SamplePower is produced by, and is available from **SPSS**.

GPower is a freeware program, available to download from the Internet at <http://www.psychologie.uni-trier.de:8000/projects/gpower.html>.

For a review of other power analysis programs, see Thomas, L. and Krebs, D. (1998, in press). A review of statistical power analysis software. *Bulletin of the Ecological Society of America*, Vol 78(2). This is also available on the **WWW**.

<!C-i''d=document;c='<img WIDTH=120 HEI



Java applets for power and sample size

Try one of these:

CI for one proportion
 Test of one proportion
 Test comparing two proportions
 CI for one mean
 One-sample t test (or paired t)
 Two-sample t test (pooled or Satterthwaite)
 Balanced ANOVA (any model)

Each selection provides a graphical interface for studying the power of one or more tests. They include sliders (convertible to entry fields) for varying parameters, plus provision for graphing one against another.

The "Balanced ANOVA" selection provides dialog with a list of several popular factorial designs, plus a provision for your own model.

Dialogs open in separate windows.

Run selection

Note: These require a web browser capable of running Java 1.1 applets. Examples include Netscape 4.5, Netscape 4.0 + an extra package, IE 4, and IE 5. If the script doesn't run, chances are that you either have disabled Java, or your browser is not new enough. Some versions of Internet Explorer incorrectly render sliders when they are in text-entry mode (call Bill G. at home and complain). The sliders still work correctly otherwise.

Advice

The most common questions I receive concern **retrospective (post hoc) power**, and using Cohen's "small", "medium", and "large" effect sizes. I strongly recommend against both of these approaches. Please see:

- Lenth, R. V. (2001), "Some Practical Guidelines for Effective Sample Size Determination," *The American Statistician*, **55**, 187-193
- Hoening, John M. and Heisey, Dennis M. (2001), "The Abuse of Power: The Prevalent Fallacy of Power Calculations for Data Analysis," *The American Statistician*, **55**, 19-24.
- An earlier draft of the Lenth reference above is [here](#), and a shorter summary of some comments I made in a panel discussion at the 2000 Joint Statistical Meetings in Indianapolis is [here](#).
- Additional brief comments, prepared as a handout for my poster presentation at the 2001 Joint Statistical Meetings in Atlanta, are [here](#).

Accuracy

- **Formula accuracy**
Most computations are "exact" in the sense that they are based on exact formulas for sample size,

power, etc. The exception is Satterthwaite approximations; see below.

- **Machine accuracy**

Even with exact formulas, computed values are inexact, as are all double-precision floating-point computations. Many computations (especially noncentral distributions) require summing one or more series, and there is a serious tradeoff between speed and accuracy. The error bound set for cdfs is 1E-8 or smaller, and for quantiles the bound is 1E-6. Actual errors can be much larger due to accumulated errors or other reasons. Quantiles, for example, are computed by numerically solving an equation involving the cdf; thus, in extreme cases, a small error in the cdf can create a large error in the quantile.

A warning (typically, "too many iterations") is generated when an error bound is not detected to have been achieved. However, in the case of quantile computations, no warning message is generated for extreme quantiles. If you want a power of .9999 at $\alpha=.0001$, you can expect the computed sample size to *not* be accurate to the nearest integer! If you specify reasonable criteria, the answers will be pretty reliable.

- **Satterthwaite approximations**

Some of the dialogs (two-sample t , mixed ANOVA) implement Satterthwaite approximations when certain combinations of inputs require an error term to be constructed. These are of course not exact, even in their formulation. Moreover, the Satterthwaite degrees of freedom is used as-is in computing power from a noncentral t or noncentral F distribution, and this introduces further errors that could be large in some cases. In the two-sample t setting, I'd expect the worst errors to exist when there is a huge imbalance in sample sizes and/or variance. In the dialogs for mixed ANOVA models (either F tests or multiple comparisons/contrasts), I expect these errors to get worse as more variance components are involved, especially when one or more of them is given negative weight. I intend to investigate this issue and see if there is room for improvement within the very general framework in which these dialogs are designed.

Java development

The file `piface.jar` may be downloaded so that you can run these applications locally. You will need a java runtime environment (JRE) which is available for free download for several platforms from [Sun](#) -- but you may have it already on your system. Put the .jar file in a place that is accessible to your JRE. You can adapt this web page, save it as an HTML file, and run it in appletviewer; or you can run a specific dialog if you know its class name -- e.g., `java TwoTGUI`. To find out the application names, view `piface.jar` with a zip file program (such as WinZip) and look at the class names in the directory `rvl/piface/apps`.

You can develop your own applications, taking advantage of what's already available in `piface.jar`. Documentation of the classes therein is available [here](#). (It is somewhat spotty; I'll try to improve it.) The main idea is to extend the class `rvl.piface.Piface`. You are only required to provide two methods: `gui()` -- which sets up the interface -- and `click()` -- which is called when something changes. `Piface` uses Java's reflection mechanism to associate variable names with user-interface components. At your option, you can write other methods with names like `variable_changed()`; if a method like this exists, then it is called when the user manipulates the control associated with `variable`. A number of user-interface elements are provided, plus a spectrum of cdfs, quantile functions, power functions, mathematical functions, etc.

Links to other sites

- A [review of power software for PCs](#) by Len Thomas and Charles Krebs

- Interactive page - [Michael Friendly](#) (ANOVA designs)
- Interactive page - [David Schoenfeld](#) (clinical trials designs; menu based on study type and measurement type)
- [UnifyPow](#) - A SAS module for sample-size analysis by Ralph O'Brien.
- [SSize](#) - ECHIP sample size calculator for Palm devices (freeware) by Bob Wheeler.

It is not unlikely that one or more of these links is broken. If that happens, please let me know (especially if you can correct it!)

This page was last modified Monday, 17-Sep-2001 11:44:58 CDT. The views and opinions expressed in this page are strictly those of the page author. The contents of this page have not been approved by the Division of Mathematical Sciences, the College of Liberal Arts or The University of Iowa.

Russ Lenth / rlenth@stat.uiowa.edu

Table 2: Verified List

Minimum number of measured exceedances needed to put on the Verified list with at least 90% confidence that the actual exceedance rate is greater than or equal to ten percent.					
Sample sizes		Are listed if they have at least this # of exceedances	Sample sizes		Are listed if they have at least this # of exceedances
From	To		From	To	
10	11	3	245	253	32
12	18	4	254	262	33
19	25	5	263	270	34
26	32	6	271	279	35
33	40	7	280	288	36
41	47	8	289	297	37
48	55	9	298	306	38
56	63	10	307	315	39
64	71	11	316	324	40
72	79	12	325	333	41
80	88	13	334	343	42
89	96	14	344	352	43
97	104	15	353	361	44
105	113	16	362	370	45
114	121	17	371	379	46
122	130	18	380	388	47
131	138	19	389	397	48
139	147	20	398	406	49
148	156	21	407	415	50
157	164	22	416	424	51
165	173	23	425	434	52
174	182	24	435	443	53
183	191	25	444	452	54
192	199	26	453	461	55
200	208	27	462	470	56
209	217	28	471	479	57
218	226	29	480	489	58
227	235	30	490	498	59
236	244	31	499	500	60

1 (1) Water segments shall be placed on the planning list if, using objective and credible
2 data, as defined by the requirements specified in this section, the number of exceedances of an
3 applicable water quality criterion due to pollutant discharges is greater than or equal to the
4 number listed in Table 1 for the given sample size. This table provides the number of
5 exceedances that indicate a minimum of a 10% exceedance frequency with a minimum of an
6 80% confidence level using a binomial distribution.