Power Analysis Overview

What is statistical power?

A hypothesis test, in the context of monitoring impact since restoration, would likely be an analysis that tests whether an indicator variable has changed over time, e.g. between your baseline values and data collected later in the funding period, and/or between your treatment group (where restoration action took place) and your control group (where no restoration action took place). Statistical power in a hypothesis test is the probability that your test will detect an effect that actually exists (i.e. a statistically significant difference). If a study has 90% statistical power, there is a 90% chance that your statistical test will detect an effect that is really there. A power analysis, therefore, computes statistical power, and is usually undertaken when you are designing your study or monitoring plan.

Statistical power is a function of three variables: sample size, effect size, and variance in the data. These four components are therefore interdependent: if you know three of the variables, you can always determine the fourth. Likewise, if you increase one variable in the equation (e.g. sample size), you will increase another variable (e.g. statistical power) to keep the equation balanced.

Why is knowing statistical power useful?

A power analysis is used either to determine the statistical power of your monitoring design, or the sample size required to detect an expected effect. In the case of determining statistical power in a study, a power analysis helps you determine if your design is over- or under-powered. An overpowered design will detect an effect that is so small it is not useful. An underpowered design may fail to detect a real effect, because the sample size is too small for the effect to be observed, given the variability in the data. In applied restoration, underpowered studies are more common, given that it is often unfeasible or too costly to increase your sample size. In an ideal world, you can avoid misuse of time and money by using a power analysis to select an appropriate sample size before your monitoring begins, providing your study with sufficient statistical power to detect expected changes in you land or seascape.

If you have a fixed sample size that cannot be adjusted, a power analysis can also help you justify if the study is worth doing or not, depending on the probability that you will detect an effect with the sample size you have.

Lastly, a power analysis may help in future grant applications, providing potential funders more confidence in funding your proposed monitoring, because you've carefully designed it to detect the expected change without overusing resources (e.g. having more samples than is required).

What information do you need to do a power analysis?

1. Knowledge of the subject of your monitoring. This can be based on expert knowledge of your restoration area or taxa of interest, a pilot study, historical data from your project area, meta-analyses, or studies conducted in other areas that are comparable to your

restoration site. Power analyses are often based on some assumptions and educated guesses, which can be refined over time.

2. The statistical test you will perform when you eventually test your hypothesis. In many cases, your hypothesis will be related to the efficacy of your restoration intervention (i.e. are there more regenerating trees in my treatment area, compared to my control area? Is there greater overall fish biomass in the marine protected area, compared to the areas where fishing is allowed?). When you know this question and the data you will collect in order to answer your question, you can determine which statistical test will be most appropriate to answer your question. Examples of common tests are one-tailed t-tests (if you would like to know if the means of one group are higher/lower than the means of another), or two-tailed t-tests (when you simply would like to know if the means between two groups are significantly different from one another).
3. Alpha, or significance level. Most power analyses will require that you set the level of significance, for which you'd like to detect a statistically significant effect in your results. This is usually set to 0.05, but a more conservative estimate of 0.01 can also be used.
4. You will need estimates of three of the four components of a power analysis. You will always need to estimate the effect size you expect to observe and the variance you expect in the data (see below for more details). Obviously, if you already knew these values, you wouldn't need to conduct the study, which is why these values are only estimates, based on educated guesses and expert knowledge (see point 1 above).

Conducting a power analysis

A power analysis is used to identify either statistical power (when sample size is provided), or sample size (when effect size is provided). In both approaches, the other two components of the analysis—estimated effect size and variance—will always need to be provided.

1. Effect size

The effect size is a standardized number, so that in statistical tests you can compare the effects of treatments with very different scales of change. Generally, effect size is calculated by taking the difference between the two groups you are comparing (e.g., the mean of treatment group minus the mean of the control group) and dividing it by the standard deviation of one of the groups (usually the control group). The larger the effect size, the less likely any observed effect will be due to random sampling error, or that small effects are masked by background variation in the data, and the easier it is for your test to detect a real effect. To interpret this, many people use this general guide developed by Jacob Cohen (Cohen, 1988): < 0.1 = trivial effect; 0.1 - 0.3 = small effect; 0.3 - 0.5 = moderate effect; > 0.5 = large effect.

Given that you have not done your study yet and you do not know the true effect size (difference between the means and standard deviation), you will have to estimate it based on your knowledge about your study subject (which can come from personal experience, published studies, a pilot study, etc.). This can be difficult to estimate, but keep in mind that it is just an estimate, and does not need to be a perfect representation of your expected data. For that reason, do not specify what you expect to find, but rather the smallest effect that would still be useful in your monitoring. What is the smallest effect that you would consider a successful change in your restoration site? What effect would be too small to be worth putting resources into? (In that case you'd want to set it higher.) You can also set your effect size based on a specific restoration goal, in which case you'd want to set your effect size slightly below that to ensure you capture it. As a last resort, you can set your effect size quite simply at .5, .3 or.1, corresponding with a high, medium or low effect (Cohen, 1988), depending on how conservatively you'd like to estimate to be.

2. Variance

There are different ways to estimate variance of a given sample, but power analyses rely on standard deviation around the mean (SD). As with effect size, the standard deviation of your dataset is unknown before you start your monitoring, so you will have to estimate this. If you have access to previous or ongoing monitoring data of your target population or habitat, you can use that variation to inform your estimate of variance for your planned monitoring. The more variability in your data, the more difficult it will be to detect a real effect, therefore the more statistical power you will need. When you take samples of a variable (e.g. tree diameter, fish biomass, ratings of people's acceptance of reintroduction of a species), do you expect there to be a wide range of collected values (high SD), or do you expect them all to be quite similar to each other (low SD)? The default standard deviation is 1, however you can increase or decrease this number if you are confident that you will have more or less variation in your data, respectively.

3. Statistical power

You will provide this value if you want to identify your optimal minimum sample size for your study. In this case, you will need to supply a numeric value between 0 and 1 that represents the probability that your statistical test will identify an effect that is really there. Again, a value set at 0.8 means your statistical test has an 80% chance of detecting a change that actually exists. If you set the value too low, you will invite too much uncertainty into your study. If you set this value too high (approaching 1), you will require an unnecessarily larger sample size. This is typically set at 0.8 or 0.9. By choosing this value carefully, you tailor the experiment so that it has a reasonable chance of detecting useful differences, while allowing smaller, non-useful differences to remain potentially undetected.

4. Sample size

You will provide this value if you want to identify the statistical power of your monitoring design. A larger sample size allows a hypothesis test to detect a smaller effect, given the other variables (statistical power and variance) remain constant. This number could be determined by the number of sites logistically available in your restoration area, or a hypothetical number of sites for which you'd like to know the statistical power.

A note on how we define samples: we are careful here to distinguish between the number of sampling sites you have available to you, and the number of samples taken at each site. The number of sites is sometimes equal to the number of samples you have; in other cases, the number of samples is greater than, and nested within, the number of sites you have. In other cases, you'll have sub-plots that you plan on averaging to get one value per macro-plot, in which case the sample size with respect to a power analysis would not be the number of sub-plots, but the number of macro-plots. In all cases, when we talk about samples, we are assuming they are independent samples. If you are unsure about how to conceptualise your sample size or their independence in your monitoring design, please feel free to consult the ELSP Science Team. This is not always a straight-forward task and can confused even very experienced researchers!

You can also supply different sample sizes into the equation, and observe how your statistical power changes. If you have 20 samples in each group, your statistical power might be 70%, meaning you have a 70% chance of detecting an effect in your study. If that is too low of a probability for you, you can re-run the test with a higher sample size, gaining a better understanding of how much statistical power you gain with every added sample, until you reach a satisfactory level of statistical power.

Tips

- Because the effect size and variance estimates are based on educated guesses, you can change these values to a range of likely outcomes, and see how much your statistical power or sample size changes. This will give you a range of scenarios that will produce a good study, in which you can then weigh the logistical trade-offs of reducing statistical power versus adding more sampling sites accordingly.
- In an effort to keep this overview brief, we will not go into how to conduct a power analysis for every statistical test, or the specific equations for each. However, power analyses can be conducted using other statistical tests, for example correlation tests, ANOVAs (comparing means of more than two groups; e.g. if you have a control group, a treatment group, and a reference group), binomial tests (where the variable contains presence-absence values or proportions, rather than continuous values), Poisson-distributed data (dealing with rates of occurrence), or multivariate models, all which will compute power differently, due to the data structure and underlying assumptions of the chosen model. The resources provided at the bottom of this document will automatically alter the equation depending on the statistical test you specify.
- Power analyses are not just used for ecological data; they can be equally useful for determining the sample size for socioeconomic indicators that you plan to analyse quantitatively (e.g. the ideal sample size for the number of participants in a survey).
- If you use a power analysis to determine the number of samples you need, please note that the number of samples provided is the number of samples per treatment group, not number of total samples. For example, if your power analysis says you need twenty samples, that is twenty samples per treatment, or category. If you have an area where you are doing your restoration action, and a similar area where you are not, you would need twenty sampling sites in each area, totalling 40 samples altogether.
- Reducing the number of treatments or groups will increase statistical power/require a smaller sample size. Do you have just a treatment and control site? Or do you have multiple treatment groups, and a control site? If you have limited resources or sampling sites, you may want to consider the trade-off of having multiple treatments, versus just one, to maintain sufficient statistical power.
- Consider a one- versus two-tailed t-test, when appropriate. If all you need to know is if you treatment variable is greater than or less than your control variable, you can use a one-tailed t-test, which increases statistical power.

- Generally, if your study employs continuous data, your required sample size is likely to be smaller than if you use binomial (presence/absence) or Poisson (rates of occurrence) data.
- If you include interactions between independent variables in your model, you will require more statistical power/a larger sample size.
- If possible, pairing your samples in your study design can increase statistical power. Paired samples means they are dependent on some additional factor (i.e. each pair of control and treatment sites are located in the same area or altitudinal gradient or protected area), whereas an unpaired design means all the samples are independent of one another (i.e. the treatment and control sites are randomly selected within your project area). In your power analysis, you would select a paired statistical test (e.g. a paired t-test), for which the power analysis will account.
- The greater total number of statistical tests that will be performed on the data set requires higher statistical power. If you are collecting multiple variables from the same sampling sites and will be performing multiple statistical tests on that dataset, best practices require a higher number of samples to account for repeated testing.

Resources

There are freely available websites for conducting a power analysis, with tutorials on how to use them, such as GigaCalculator.com, WebPower, or GPower, among others. The interfaces are slightly difference, but all require the same information.

If you a familiar with R, you can conduct a power analysis using the 'pwr', 'TrialSize', or 'PowerUpR' packages.

If you would like to read more on the topic, there are a number of excellent websites, tutorials and books on the topic, a few of which are listed here:

Colgrave & Ruxton (2021). Power Analysis: An Introduction for the Life Sciences (Oxford Biology Primers). Oxford University Press: Oxford, UK.

Power Analysis in R (ladal.edu.au)

A Gentle Introduction to Statistical Power and Power Analysis in Python - MachineLearningMastery.com

Introduction to Power Analysis (ucla.edu)

References

Cohen, J. (1988). Statistical Power Analysis for the Behavioral Sciences, Second Edition. Hillsdale, New Jersey: Lawrence Erlbaum Associates.