

When does it make sense to use statistical analysis?

First, before we get into tips about when does it make sense to use statistical analysis, let's define what it is.

Definition: Statistical analysis is the manipulation, summarization, and interpretation of quantitative data.

So, basically whenever you want to take a look at the differences your programs make for the participants you serve it makes sense to do some type of statistical analysis. In our case we want to look at the change in outcomes (knowledge, attitudes, behaviors, skills and status) of our participants to determine if our programs are achieving the results they set out to achieve.

For the most part you will be reporting the outcomes of your program activities based on your logic model and the benchmarks you set.

Question: *Did we reach our desired outcomes based on the change in the pre-post test scores or other form of evaluation that we used?*

Reporting whether or not you reached your outcome benchmark for your activities might be enough as far as your reporting requirements go, but there may be situations where you want to look a little more deeply at your program data. For instance, you may want to see if changing the frequency and duration of your activities makes a difference in your outcomes. Or, you may notice some difference in outcomes based on the staff conducting the activities or the age group of the participants that you want to investigate further.

If you want to take on this investigation on your own, then the following information will help you do that. If you feel overwhelmed already, then its time to turn to someone to help you....*read no more.....*

So if you are still reading, then you think you want to try this on your own OR you want to know the nuts and bolts so you can understand what the outside help that you are working with is going to need.

Question: *Do you have enough cases that include a pre and a post test to analyze?*

A general rule is that you should have at least 75 cases that include a matched pre and a post test (or some type of "before" measurement and some type of "after" measurement). Now, if you want to look at your results for specific groups (such as grade levels) you will also need to have at least 75 students in each group for both the pre and post test scores. It is even better if you have at least 100 cases that include a matched pre and a post test. This is based on the assumption that you will be conducting an analysis to compare the pre-test "score" with the post-test "score".

The question then becomes: *When does it make sense to use what type of statistical analysis?*

Ask yourself: *What is my research question...do I want to compare between pre-program and post-program, do I want to know if there are specific factors that influence the differences in change scores....think carefully about your research question(s).*

The choice of a data analysis method is affected by several considerations. These include:

- the level of measurement for the variables to be studied;
- the unit of analysis;
- the shape of the distribution of a variable, including the presence of outliers (extreme values);

- the completeness of the data;
- the number of cases you have, and;
- most importantly the evaluation question(s) you want to answer.

First, knowing the level of measurement helps you decide how to interpret the data from that variable. When you know that a measure is nominal, then you know that the numerical values are just short codes for the longer names. Second, knowing the level of measurement helps you decide what statistical analysis is appropriate on the values that were assigned. If a measure is nominal, then you know that you would never average the data values or do a t-test on the data.

Levels of measurement

There are four types of levels of measurement. Often, the last two levels of measurement are grouped together in analysis even though you may use different types of analysis on the last two levels.

Level of Measurement	Description	Example
Nominal	The attributes of a nominal variable have no inherent order. For purposes of data analysis, we can assign numbers to the attributes of a nominal variable but must remember that the numbers are just labels and must not be interpreted as conveying the order of the attributes.	Gender is a nominal variable in that being male is neither better nor worse than being female. Persons, things, & events characterized by a nominal variable are not ranked or ordered by the variable.
Ordinal	With an ordinal variable, the attributes are ordered. Although the ordinal level of measurement yields a ranking of attributes, no assumptions are made about the “distance” between the classifications. For data analysis, numbers are assigned to the attributes (for example, greatly dislike = -2, moderately dislike = -1, indifferent to = 0, moderately like = +1, and greatly like = +2), but the numbers are understood to indicate rank order and the “distance” between the numbers has no meaning.	Observations about attitudes are often arrayed into five classifications, such as greatly dislike, moderately dislike, indifferent to, moderately like, greatly like.
Interval	The attributes of an interval variable are assumed to be equally spaced. However, it is not assumed that a 90-degree object has twice the temperature of a 45-degree object (meaning that the ratio of temperatures is not necessarily 2 to 1). The condition that makes the ratio of two observations uninterpretable is the absence of a true zero for the variable. In general, with variables measured at the interval level, it makes no sense to try to interpret the ratio of two observations.	Temperature on the Fahrenheit scale is an interval variable. The difference between a temperature of 45 degrees and 46 degrees is taken to be the same as the difference between 90 degrees and 91 degrees.
Ratio	The attributes of a ratio variable are assumed to have equal intervals and a true zero point. With ratio variables, it makes sense to form ratios of observations and it is thus meaningful, for example, to say that a person of 90 years is twice as old as one of 45. For analysis purposes, it is seldom necessary to distinguish between interval and ratio variables so we usually lump them together and call them interval-ratio variables.	Age is a ratio variable because the negative age of a person or object is not meaningful and, thus, the birth of the person or the creation of the object is a true zero point. Pre/Post-test scores can be considered ratio if there is the option of the individual to score a “0” on the pre/post test.

Unit of analysis

Units of analysis are the persons, things, or events under study—the entities that we want to say something about. Frequently, the appropriate units of analysis are easy to select. They follow from the purpose of the study. For example, if we want to know how people feel about the primary prevention of sexual violence, individual people would be the logical unit of analysis. In the statistical analysis, the set of data to be manipulated would be variables defined at the level of the individual.

However, in some studies, variables can potentially be analyzed at two or more levels of aggregation. Suppose, for example, that evaluators wished to evaluate the Safe Dates program and had acquired pre and post test scores on a large number of students, some who participated in the program and some who did not. One way to analyze the data would be to treat each child as a case.

But another possibility would be to aggregate the scores of the individual students to the classroom level. For example, they could compute the average scores for the students in each classroom that participated in their study. They could then treat each classroom as a unit, and an average pretest and post-test score would be an attribute of a classroom. Other variables, such as presenter's years of experience, number of the classroom level. The data analysis would proceed by using classrooms as the unit of analysis. For some issues, treating each student as a unit might seem more appropriate, while in others each classroom might seem a better choice. And we can imagine rationales for aggregating to the school, school district and even state level.

In summary, the unit of analysis is the level at which analysis is conducted. We have, in this example, five possible units of analysis: student, classroom, school, school district, and state. We can move up the ladder of aggregation by computing average pretest/posttest scores across lower-level units. In effect, the definition of the variable changes as we change the unit of analysis. The lowest-level variable might be called student-pretest/posttest-score, the next could be classroom-average-pre-test/post-test-score, and so on.

In general, the results from an analysis will vary, depending upon the unit of analysis. Thus, for studies in which aggregation is a possibility, evaluators must answer the question: ***What is the appropriate unit of analysis?*** Several situation-specific factors may need consideration, and there may not be a clear-cut answer. Sometimes analyses are carried out with several units of analysis.

Distribution of a variable

The cases we observe vary in the characteristics of interest to us. For example, students vary by class and by pre/post-test scores. Such variation across cases, which is called the distribution of a variable, is the focus of attention in a statistical analysis. For example, some techniques are suitable only when the distribution is approximately symmetrical, while others can be used when the observations are asymmetrical. Once data are collected for a study, we need to inspect the distributions of the variables to see what initial steps are appropriate for the data analysis. Sometimes it is advisable to transform a variable (that is, systematically change the values of the observations) that is distributed asymmetrically to one that is symmetric. Another aspect of a distribution is the possible presence of outliers, a few observations that have extremely large or small values (or scores on their pre/post tests) so that they lie on the outer reaches of the distribution.

Completeness of the data

When we design an evaluation, we plan to obtain data for a specific number of cases. Despite our best plans, we usually cannot obtain data on all variables for all cases. For example, in a pre/post survey, some persons may decline to respond at all and others may skip certain items. Or responses to some questions may be inadvertently

“lost” during data editing and processing. Almost inevitably, the data will be incomplete in several respects, and data analysis must contend with that eventuality.

Incompleteness in the data can affect analysis in a variety of ways. The classic example is when we draw a probability sample with the aim of using inferential statistics to answer questions about a population. To illustrate, suppose you request that a teacher administers the pre/post test to all of his/her students, but only 45 percent of the teachers that you ask provide data. Without increasing the response rate or satisfying themselves that non-respondents would have answered in ways similar to respondents (or that the differences would have been inconsequential), you would not be entitled to draw inferential conclusions about the student population that attend your program. You can write this up in your results section as a “caveat” or limitation to your evaluation.

The methods of statistical analysis provide us with ways to compute and interpret useful statistics. Those that are useful for describing a population are called descriptive statistics. They are used to describe a set of cases upon which observations were made.

Methods that are useful for drawing inferences about a population from a sample of cases are called inferential statistics. They are used to describe a population using merely information from observations on a sample of cases from the population. Thus, the same statistic can be descriptive or inferential or both, depending on its use.

Descriptive statistics

A descriptive statistic is a number, computed from observations of a group, that in some way describes the group of cases. The definition of a particular descriptive statistic is specific, sometimes given as a recipe for the calculation. Measures of central tendency form a class of descriptive statistics each member of which characterizes, in some sense, the typical value of a variable—the central location of a distribution.

Measures of Central Tendency

Measurement Level	Use of Measure		
	Mode	Median	Mean
Nominal	Yes	No	No
Ordinal	Yes	Yes	No
Interval	Yes	Yes	Yes
Ratio	Yes	Yes	Yes

Spread refers to the extent of variation among cases—sometimes cases cluster closely together and sometimes they are widely spread out. When we determine appropriate policy action, the spread of a distribution may be as much a factor, or more, than the central tendency.

Measures of Spread

Measurement Level	Use of Measure			
	Index of Dispersion	Range	Interquartile Range	Standard Deviation
Nominal	Yes	No	No	No
Ordinal	Sometimes	Sometimes	Yes	No
Interval	No	Yes	Yes	Yes
Ratio	No	Yes	Yes	Yes

Note: The index of dispersion is a measure of spread for nominal or ordinal variables. With such variables, each case falls into one of a number of categories. The index shows the extent to which cases are bunched up in one or a few categories rather than being well spread out among the available categories.

See the tip sheet “Different ways to present quantitative evaluation results” for additional information about descriptive statistics.

Inferential statistics

With inferential statistics, you are trying to reach conclusions that extend beyond the immediate data alone. For instance, we use inferential statistics to try to infer from the sample data what the population might think. Or, we

use inferential statistics to make judgments of the probability that an observed difference between groups is a dependable one or one that might have happened by chance in this study. Thus, we use inferential statistics to draw conclusions from our data to more general conditions.

Types of evaluation questions you might ask and the statistical method to use (T-tests, ANOVA and/or Chi Square)

Evaluation Question Scenario 1: T-test

Whenever you wish to compare the average performance between two groups you should consider using the t-test. Perhaps your question is: *Do the boys participating in the program score differently than the girls participating in the program?* The t-test will analyze the scores and let you know if there is a statistically significant difference between the scores of the boys and girls.

Evaluation Question Scenario 2: Analysis of Variance (ANOVA)

So you are wondering if you really need to work with the participants for 3 hours a day 2 days a week or if you could perhaps change your time with the students to 3 days a week for 1 hour each time. Perhaps the evaluation question is: *Does the dose of the program make a difference to the program outcomes? In this case is the 6 hour program more effective than the 3 hours of program?* If you have the scores for a large enough sample of participants in both program situations then you can compare the effectiveness of the dosage by using an ANOVA.

Evaluation Question Scenario 3: Chi-Square

So you have a group of surveys that ask questions about attitudes toward sexual harassment and you want to know if males are more likely than females to support attitudes about sexual harassment. Evaluation question: *Do the male students more often hold attitudes that support sexual harassment than the female students in the schools we work in?* You can use the data from your likert-type scale to develop collapsed categories of attitudes about sexual harassment (Pro/Con). Because you basically have “nominal” data (categories without true values), you can look at the data using a chi-square analysis. The chi-square analysis will tell whether or not the proportion of females that support attitudes toward sexual harassment is higher than the males (or vice versa). You could also expand this question by using the chi-square to look at differences between male/female and also grade level.

For more information

One hour video on “the joy of stats” by Hans Rosling

<http://www.gapminder.org/videos/the-joy-of-stats/>

Research Methods Knowledge Base, Web Center for Social Research Methods

http://www.socialresearchmethods.net/kb/stat_t.php