

Basic Statistics for Evaluation

A Necessary Language

Numerical results provide a way to describe positive aspects of our work. In this way, statistics are a necessary part of program evaluation and outcomes analysis. The webinars and resources on this site will help you realize that basic statistical methods are approachable and manageable.

Basic statistics help us communicate our program results and also understand the findings of our peers, in their articles and presentations. Some terms seem foreign, but a brief review will make them more familiar. We already speak the language of statistics more than we realize, for example when we talk about the likelihood that it will rain, or when we anticipate a certain result.

To review some basic terms of statistics see, for example, Edward Cohen's webinar [Performance Measurement – Using Data to Improve Quality](#) and Greg Vinson's webinar [Demonstrating Client Improvement to Yourself and Others: Setting up an Evaluation System to Succeed](#). They provide basic terms that apply to the context of evaluation.

Making Inferences

Most of our Evaluation processes involve inferential and descriptive statistics.

Inferential statistics uses information from a specific sample to make general statements (predictions, and inferences) about the population at large. If our sample represents a larger population, then analysis of our sample tells us about the population from which that sample is a subset. In our community of research, we have no resource for the overall population of torture survivors: there is no census, no national database. We work from samples, from information about our own client bases. We already have a sense that our observations apply to other settings, but inferential statistics helps us say, with a level of certainty and confidence, that something we see in our own client base applies to the community of survivors at large. Statistics are a necessary step, to report results with confidence.

Descriptive statistics provides ways we can show, demonstrate, or describe the results of our work. The webinar [Presenting Outcomes Data Clearly and Effectively](#) reviews ways to present your findings, and includes a template you can modify to create your own graphs.

Measuring change

For most evaluation and outcomes analysis, the most basic and useful analysis is to compare measurements taken at two points in time. A meaningful measure will be a mean, or average of the values collected. For example, to examine PTSD scores for our clients, at two points in time, we could use a t-test to see if there is any significant difference between the two average values. An Anova will compare two or more means simultaneously, and also takes into consideration the variance within a sample and between samples. The examples within these webinars use t-tests. Your own initial evaluation and outcomes analysis will only require a t-test.

In his webinar [Demonstrating Client Improvement to Yourself and Others: Understanding and Using your Outcome Evaluation System](#) Greg Vinson shows how you can use Excel to set up and perform a t-test on your data, to identify whether there is a difference between two means. The results of your t-test will confirm whether there was a change, whether your intervention had an effect.

Making sense of results

To decipher the results of a t-test, examine the resulting p-value in your Excel output. The p-value identifies the probability that the result occurred by chance. We do not want results to occur by chance; we want our result to occur because of our intervention. We therefore want the p-value to be small, less than .05 in general. If your p-value is less than .05, this means there is less than a 5% chance that your results occurred by random chance. The range ($p < .05$) is a standard for most research, and an acceptable level of random chance.

This single number (the p value), generated by one additional step you can do within Excel will allow you to state your results with confidence. If you find that your data yields a small p value, then you can state with a certain level of confidence that your intervention led to the results you observe. The webinar shows how you can [report your findings in APA format](#).

Once you determine whether your program has an effect, an impact on your clients' lives, you can see whether the effect was small, medium, or large. To check the size of the effect, we calculate a Cohen's d value. The webinar shows step by step, how to generate a Cohen's d value and how to interpret it.

Trusting your results

Once we evaluate a measure, we also need to review the validity and reliability of our results. We can use statistical methods to examine reliability and validity in order to minimize false conclusions.

In his webinar [Assessing Trauma & Associated Symptoms in Refugees & Torture Survivors](#), Dr. Hollifield notes the need to use measurement instruments that have already been tested for reliability and validity for your study population. An analysis of reliability and validity show that are results are accurate and meaningful. In his webinar, [Demonstrating Client Improvement to Yourself and Others: Making Sure Client Numbers Reflect Client Reality](#), Greg Vinson shows the steps to evaluate reliability and validity.

The webinars and resources on this site provide the basic guidance you need to take these steps to collect your own data, review your results, and report them with statistical certainty. You are encouraged to do your own data collection and analysis. You will be empowered to speak the language of numbers.

This summary is based on information in a series of National Capacity Building Institute webinars (presented 2007-2012). The content of this summary is based on the insight, expertise, and findings of several presenters: Greg Vinson, Michael Hollifield, Ken Miller, Edward Cohen, and Joan Othieno.