Roadmap to Data Analysis

Introduction to the Series, and

I. Introduction to Statistical Thinking-A (Very) Short Introductory Course for Agencies

Objectives of the Series "Roadmap to Data Analysis"

- Provide and introduction to basic statistical procedures relevant to SOT agencies
- Provide a foundation to analyze and interpret data on clients, services, and outcomes
- Provide an introduction to the use of available tools for basic analysis of data
- Provide an understanding of the limitations of statistical analysis, and guidelines for agency staff about when to seek statistical consultation

The Series “Roadmap to Data Analysis”

I. Introduction to Statistical Thinking
II. Primer on Measurement and Variables
III. Choosing the Right Statistical Test
IV. Comparing Averages – The *t* Tests
V. Comparing Counts – Chi Square
VI. Relationship Between Two Continuous Variables – Pearson’s *r* Correlation

Learning Objectives:
I. Introduction to Statistical Thinking

- Understand basics of statistical thinking and inference
- Understand concepts of “population” and “sample” in quantitative research
- Understand the process by which inferences can be made to a population based on a sample
- Understand “hypothesis testing” and “probability value”

Inference

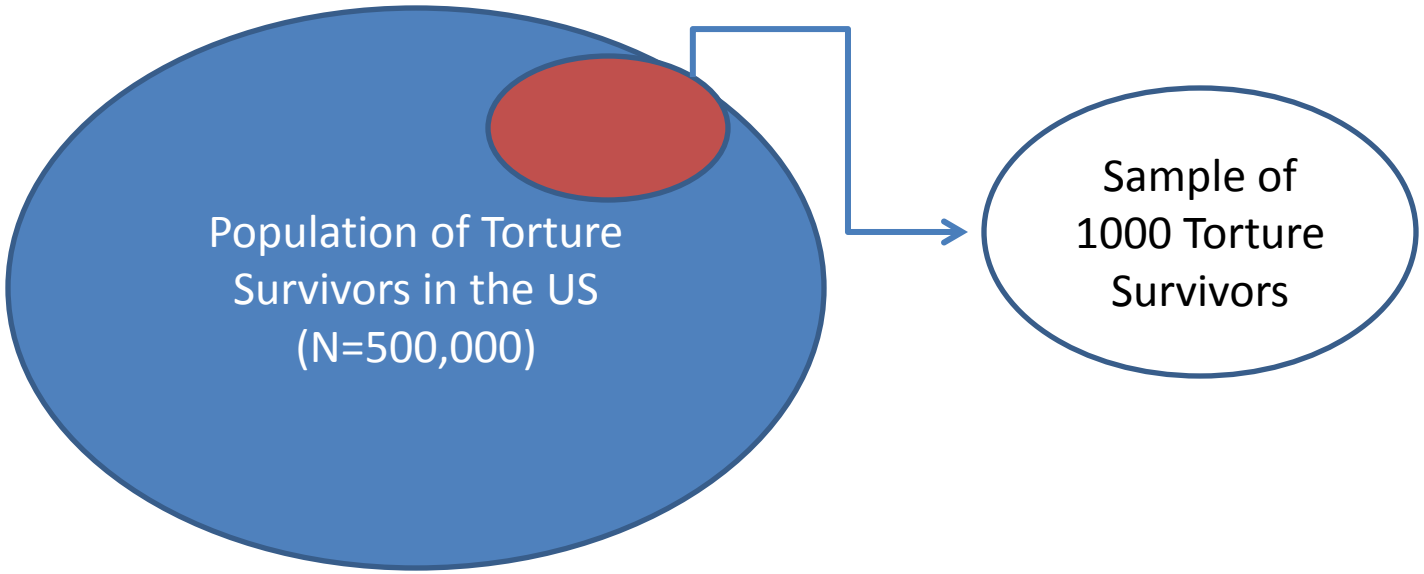
- Statistical analysis is all about understanding ("making inferences" about) a population based on a sample from the population.
  - We rarely have the opportunity to measure an entire population, such as all torture survivors in the U.S.
  - Your agency has ready-made "samples" of the population of torture survivors
- The extent to which a sample is representative of the population can be quantified
  - By a statistic – some measure of the sample (i.e. an average), and
  - By a quantified value that explains how well your sample statistic describes the population (i.e. a standard deviation)

Definitions:

- "Population": *the entire universe of individuals to which you seek to generalize*
- "Sample": *that part of the population from which you collect data*
  - A sample should be representative of the population from which it is drawn
  - The extent to which a sample is representative of the population can be quantified. This forms the basis for all types of statistical inference

Characteristics of the Sample

- A sample should be representative of the population from which it is drawn



- Note: you can also define your population as all of your agency's clients. If you take a sample from that population, then you are inferring statistical results to the agency population.

A practical example

- If we say “the longer torture survivors in your agency participate in socialization groups, the more they will show improved functioning,” *we are implying that, in the general population of torture survivors, more participation in socialization groups will result in improved functioning.*
- How is this type of inference possible?

Minimum requirements:

- The sample is reasonably representative of the population to which it will generalize
- There is a specific research question that addresses the sample and can be answered quantitatively
  - "Is the length of participation in socialization groups related to improved functioning?"
- There is an hypothesis implied by the research question that can be tested
- The analysis of data correctly matches the type of data being analyzed (i.e. using the right statistic for the data)

Hypothesis testing

- An hypothesis is *a statement about the relationship among variables*
- A variable is a *factor, or characteristic, that varies*
- "It is hypothesized that the longer torture survivors participate in our agency's socialization group, the better the functioning ."
  - Two variables
    1) Length of participation in the group, and
    2) a measure of functioning
- Hypotheses should be generated based on the available scientific literature and some theoretical basis (i.e. what leads you to think the length of participation in socialization groups might improve functioning?)

Statistics for inference

- Analysis of the data results in a statistic about the sample
  - A number representing improved functioning, such as the difference between the average pre- and post-functioning scores
  - A quantified value that explains how well your sample statistic describes the population (such as a standard deviation—a measure of how much individuals in the sample "deviate" from the average), and
  - A quantified value of how confident you can be that the sample statistic reflects the population (a "probability value")

Probability Value

- A "*p* value" *is the mathematical probability that a relationship between variables found within a sample may have been produced by chance or error*
- *P* < .05
  - "Statistically significant at <.05" – this is a typical threshold
    - Technical note: *p* values are calculated based on distribution tables for the specific statistic used. The *p* value is provided by statistics software programs
  - To interpret: "the probability is less than 5 in 100 that improvement in functioning was the result of chance alone."
  - In other words, the improvement in functioning from your sample likely reflects the same experience in the larger population – that length of participation in the socialization group is related to improved functioning (putting aside the thorny issue of cause and effect, for now…)

Important Caveats (Partial List!)

- *P* values are useful but can't tell the whole story about treatment effects (look up "effect size"). Also, *non-significant findings* can be valuable information as well.
- Statistics can't compensate for poorly written questions or instruments that are not valid
  - Especially in torture treatment settings, there are too few instruments that have been tested with many of our current survivor groups
- Statistics can't compensate for a weak research design – i.e. for understanding the impact of treatment it's best to have a non-treatment control or comparison group (to address that thorny causality issue)
- Seek consultation from statistical and content experts when considering or implementing data collection instruments
  - They can comment on reliability and validity concerns
  - They can guide analysis strategies, including sample size requirements and choosing the right statistical analyses
  - They can help interpret results