Statistical Tools To Produce Accurate and High-Value DEI Insights

Dr. Steven Sherrin, Wentworth Institute of Technology

Abstract

(Author note: I am looking for feedback on a planned workshop that focuses on improving statistical inferences with DEI data. Below is the description of that workshop.)

This session will equip participants with new research ideas and statistical tools for analyzing, visualizing, and communicating diversity, equity, and inclusion (DEI) data in engineering disciplines. The session will focus on popular, free software, and will use statistical tools to address real-world examples. Example code and visualization will be provided, as well.

The session will be organized by four core DEI principles that can help institutions create high-value, accurate, and unbiased data insights:
1. Respect differences between groups, but don't overestimate them.
2. Respect similarities between group members, but don't stereotype them.
3. Account for the multifaceted nature of individual identity.
4. Analyze small sample sizes, but don't overstate their importance.

More details on each section are provided below:

• I will demonstrate how some summary statistics, such as the mean or median, can cause readers to overestimate demographic differences. I will show an easy-to-calculate alternative (effect sizes) that counteracts this bias.

• I will demonstrate how visualizations focusing on differences between demographic groups can lead stakeholders to underestimate variation within groups. I will present a statistical technique (cluster analysis) that naturally describes within-group diversity. In addition, I will provide a simple data visualization technique, outcome-based categorization, that can also be helpful.

• I will illustrate how demographic categories commonly used by higher education institutions can fail to represent the rich, multifaceted nature of individual identity(s). I will discuss examples of how to integrate standard demographic categories with meaningful information from other datasets, such as hometown information, family migration history, and more. I will provide links and example datasets to aid audience members in conducting their own analyses.

• I will discuss two statistical biases related to the analysis of small sample sizes: 1) trying to detect meaningful patterns where there are none and 2) being overconfident in the power of small sample sizes. I will show how researchers can use a simple statistical tool, sample size calculators, to help their stakeholders avoid these errors.