1. INTRODUCTION

Equivalence Testing (ET) is a technique used to compare effect sizes or means of two studies to ascertain if they would be statistically equivalent.

OBJECTIVE

to teach Equivalence Testing applied to Educational Research, recommending emphases aimed at increasing research quality.

How does ET work?

Two One-Sided Tests (TOST)

HA1 and HA2: two alternative hypotheses
H01 and H02: two null hypotheses
ES: Effect Size
ΔL: Lower Equivalence Limit
ΔU: Upper Equivalence Limit
p-value: probability of observing a test statistic's value or smaller in a distribution, given a true null hypothesis
α: Alpha level

ET can be a useful tool for comparing means and effects within certain bounds that could hopefully imply a practical significance to provide meaning to findings.

Comparing Effect Sizes and their Confidence Intervals: A Primer on Equivalence Testing in Educational Research

1. One Single Sample Test after a Pre- and Post-Test

Data and Equivalence Bounds taken from De Muth (2019). Objective: to compare the findings (post-test scores) with a target value (Population's mean = 100 points).

This example contextualized ET in educational research with an experimental design. The problem was solved:

• through manual calculations using Critical Values of the t distribution table.
• using an online calculator to estimate p values from t values and SPSS 24 for the NHSST.

Hypotheses:
Target value = μ
Sample's mean = μ0

Conclusion:
If the null hypotheses were rejected because p values < alpha, then ET is shown to have statistical equivalence with a 90% confidence interval between the sample and population means.

Conclusion:
if the null hypotheses were not rejected because p values ≥ alpha, then ET is shown to have statistical equivalence with a 90% confidence interval between the sample and population means.

Sample's CI (Confidence Interval) 95% mean difference [-0.855, 0.555], which represented a population, was within ΔL (-1.5) and ΔU (+1.5), so it can be considered statistically equivalent to the target of 100 points.

2. Two Independent Group's Tests After a Pre- and Post-Test

Data from De Muth (2019). Scenario: Researcher A was trying to replicate a study carried out by Researcher B to observe if a treatment to improve reading had the same significance in both studies. Research question: Does statistically significant equivalence exist between these two means?

Example using:
• SPSS 24 and manual calculations with Critical Values of the t distribution table and online calculators.
• R code created by Lakens (2017) for Equivalence Testing.

Hypotheses:
If the Nulls are rejected; equivalence is proven.
Conclusion:
given a target of 100, the lower limit was set at -2 and the upper one at +2. The null hypothesis (H01) was not rejected because of the CI [-2.4638, 0.3638], which exceeded the range of |2|. Furthermore, the t-ratio failed to reject the hypothesis for the lower limit t(9) = 1.22, p = .25. Thus, there was a failure to show equivalence.

3. CONCLUSIONS

ET can be a useful tool for comparing means and effects within certain bounds that could hopefully imply a practical significance to provide meaning to findings.

REFERENCES: