Introduction

Statistical significance means a *p*-value less than some chosen α (e.g., 0.05).
We tend to think statistical significance implies there is something more than random sampling error.
Small *p*-values suggest results have meaning within the study context.
This is usually true but sometimes false.
Statistical significance is not the same as **practical significance**.

Sample Size Influences *p*-Values

One challenge:the *p*-value is related to sample size.
A small difference between the hypothesized mean (μ_0) and sample mean ($\bar{X}$) can yield a small *p*-value, even if the difference is *practically*insignificant.
A large sample size (*n*) can cause this erroneous conclusion.
Conversely, a small *n* can cause the opposite, erroneous conclusion.
A large difference between μ_0 and $\bar{X}$ can still yield a large *p*-value if *n*is small.
If *n* is too small, you might not have collected enough data to show the difference between the sample statistic and the hypothesized population parameter.

Practical Significance

One aspect of practical significance is **effect size**.
Effect size relates to the practical value of the effect.
Example: emotional intelligence study
$\mu_0 = 120$
$n = 130$
$\bar{X} = 122$
The p-value is very close to zero.
Difference between $\bar{X}$ (122) and μ_0 (120) is just 2 points.
Is this practically significant? What does this mean in terms of EI?
If not practically significant, statistical significance doesn't matter.

A Practical Significance Benchmark

What makes a value practically significant?
For emotional intelligence scores, we might review the literature:
Imagine that research says a 5-point EI difference can affect a person's ability to function emotionally.
This provides a benchmark for practical significance.
The benchmark tells us that our 2-point difference isn't meaningful, regardless of the small p-value.

Effect Size and Cohen's *d*

Effect size is a way to measure practical significance.

Cohen's *d* is a statistical method that allows us to calculate an effect size.

$$d = \left| \frac{(\bar{x} - \mu_0)}{s} \right|$$

For the EI example:

$$d = \left| \frac{(122 - 120)}{15} \right| = 0.133$$

Cohen provided a framework to categorize effects as small, medium, or large.

$d \leq 0.2$ signifies a "small" effect.

Our $d = 0.133$ is below that threshold, so it is a "small" effect.

Any statistical significance might be the result of a large n.

$d > 0.8$ signifies a "large" effect.

Even without statistical significance, $d \geq 0.8$ suggests a large effect.

$0.2 < d < 0.8$ signifies a "medium" effect.

This doesn't clearly suggest whether an effect is practically significant.

Summary

A statistically significant p-value does not imply useful results.

We want useful results to help us make decisions about practical issues.