6.4 - Practical Significance

You have previously learned how to identify statistically significant differences using null hypothesis significance testing. If the *p* value is less than the α level (typically $p < 0.05$), then the results are considered to be **statistically significant**. Results are said to be statistically significant when the difference between the hypothesized population parameter and observed sample statistic is large enough to conclude that it is unlikely to have occurred by chance.

Practical significance refers to the magnitude of the difference, which is known as the **effect size**. Results are practically significant when the difference is large enough to be meaningful in real life. What is meaningful may be subjective and may depend on the context.

Note that statistical significance is directly impacted by sample size. Recall that there is an inverse relationship between sample size and the standard error (i.e., standard deviation of the sampling distribution). Very small differences will be statistically significant with a very large sample size. Thus, when results are statistically significant it is important to also examine practical significance. Practical significance is not directly influenced by sample size.

Effect Size

For some tests there are commonly used measures of effect size. For example, when comparing the difference in two means we often compute Cohen's d which is the difference between the two observed sample means in standard deviation units:

$$d = \frac{\overline{x}_1 - \overline{x}_2}{s_p}$$

Where s_p is the pooled standard deviation

$$s_p = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}$$

Below are commonly used standards when interpreting Cohen's d:

Cohen's d	Interpretation
0.000 - 0.199	Little or no effect
0.200 - 0.499	Small effect size
0.500 - 0.799	Medium effect size
0.800 - 1.399	Large effect size
1.400 and above	Very large effect size

Example: SAT Scores

Research question: Are SAT-Math scores at College A better than at College B?

H_0: College A = College B

H_a: College A > College B

Data are collected from a random sample of 2,500 students at College A and 2,500 students from College B. In the College A sample, their average SAT Math score was 506 points, and in College B it was 500 points. For both samples, their standard deviation was 100 points.

A t-test provides a p-value of p=0.0339; therefore, the null hypothesis should be rejected, and these results would be considered statistically significant. But, let's also consider practical significance. The difference between a SAT-Math score 500 and a SAT-Math score of 506 is only 6 points; that's a very small difference.

When we compute Cohen's *d* we see that the effect size is only 0.060. $\frac{506-500}{100} = 0.060$

In most cases, this effect would **not** be considered practically significant.

Example: Commute Times

Research question: Are the mean commute times different in Atlanta and St. Louis?

Descriptive Statistics: Commute Time

City	N	Mean	StDev
Atlanta	500	29.110	20.718
St. Louis	500	21.970	14.232

Using the dataset built in to StatKey, a two-tailed randomization test was conducted resulting in a *p* value < 0.001. Because the null hypothesis was rejected, the results are said to be statistically significant.

Practical significance can be examined by computing Cohen's *d*. We'll use the standard equation:

$$d = \frac{\overline{x}_1 - \overline{x}_2}{s_p}$$

Where s_p is the pooled standard deviation

$$s_p = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}$$

Because the mean compute time for Atlanta was 29.110 minutes and the mean commute time for St. Louis was 20.718 minutes, and the pooled standard deviation was 17.773:

$$d = \frac{29.110 - 21.970}{17.773}$$

The above equation simplifies to a Cohen's d value of 0.402:

$$d = \frac{7.14}{17.773}$$

$$d = 0.402$$

The mean commute time in Atlanta was 0.402 standard deviations greater than the mean commute time in St. Louis. Using the guidelines for interpreting Cohen's *d* in the table above, this is a small effect size, despite the p-value being quite low ($p < .001$). Therefore, while the difference between the commute times in the two cities was statistically significant, its practical significance is more limited.