HYPOTHESIS TESTING

A technique to help determine whether a specific treatment has an effect on the individuals in a population.

Goal

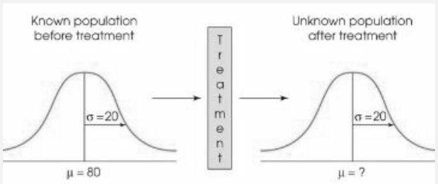
To rule out chance (sampling error) as a plausible explanation for the results from a research study.

Usage:

To evaluate the results from a research study in which:

1. A sample is selected from the population.
2. The treatment is administered to the sample.
3. After treatment, the individuals in the sample are measured.

HYPOTHESIS TESTING



If the individuals in the sample are noticeably different from the individuals in the original population, we have evidence that the treatment has an effect. However, it is also possible that the difference between the sample and the population is simply sampling error

HYPOTHESIS TESTING

Purpose:

1. The difference between the sample and the population can be explained by sampling error (there does not appear to be a treatment effect)
2. The difference between the sample and the population is too large to be explained by sampling error (there does appear to be a treatment effect).

ERRORS

TYPE I ERRORS

Occurs when the sample data appear to show a treatment effect when, in fact, there is none. In this case the researcher will reject the null hypothesis and falsely conclude that the treatment has an effect.

These are caused by unusual, unrepresentative samples. Just by chance the researcher selects an extreme sample with the result that the sample falls in the critical region even though the treatment has no effect.

The hypothesis test is structured so that Type I errors are very unlikely; specifically, the probability of a Type I error is equal to the alpha level.

ERRORS

TYPE II

Occurs when the sample does not appear to have been affected by the treatment when, in fact, the treatment does have an effect. In this case, the researcher will fail to reject the null hypothesis and falsely conclude that the treatment does not have an effect.

Type II errors are commonly the result of a very small treatment effect. Although the treatment does have an effect, it is not large enough to show up in the research study.

ERRORS

		Actual Situation	
		No Effect, H_0 True	Effect Exists, H_0 False
EXPERIMENTER'S DECISION	Reject H_0	Type I error	Decision correct
	Retain H_0	Decision correct	Type II error

Actual Solution

MEASURING EFFECT SIZE

A hypothesis test evaluates the *statistical significance* of the results from a research study. That is, the test determines whether or not it is likely that the obtained sample mean occurred without any contribution from a treatment effect.

The hypothesis test is influenced not only by the size of the treatment effect but also by the size of the sample. Thus, even a very small effect can be significant if it is observed in a very large sample. Because a significant effect does not necessarily mean a large effect, it is recommended that the hypothesis test be accompanied by a measure of the **effect size.**

We use **Cohen=s d** as a standardized measure of effect size. Much like a z-score, **Cohen=s d** measures the size of the mean difference in terms of the standard deviation.

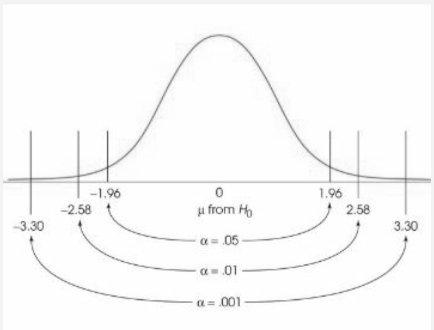
NULL HYPOTHESIS

Step 1

State the hypotheses and select an α level. The **null hypothesis**, H0, always states that the treatment has no effect (no change, no difference).

According to the null hypothesis, the population mean after treatment is the same is it was before treatment. The α level establishes a criterion, or "cut-off", for making a decision about the null hypothesis. The alpha level also determines the risk of a Type I error.

NULL HYPOTHESIS

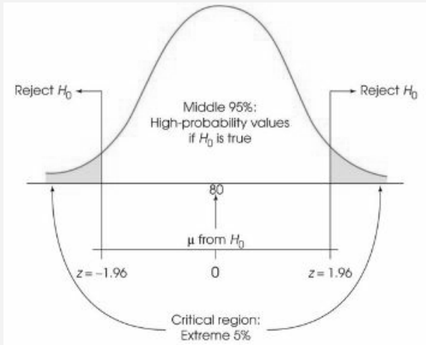


CRITICAL REGION

Step 2

Locate the **critical region**. The critical region consists of outcomes that are very unlikely to occur if the null hypothesis is true. That is, the critical region is defined by sample means that are almost impossible to obtain if the treatment has no effect. The phrase *"almost impossible"* means that these samples have a probability (p) that is less than the alpha level.

CRITICAL REGION



TEST STATISTIC

Step 3

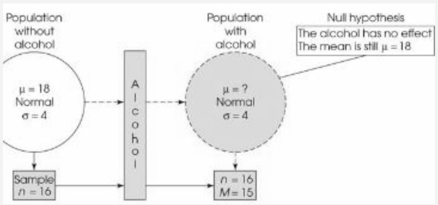
Compute the **test statistic**. The test statistic (in this chapter a z-score) forms a ratio comparing the obtained difference between the sample mean and the hypothesized population mean versus the amount of difference we would expect without any treatment effect (the standard error).

ALPHA LEVEL

Step 4

A large value for the test statistic shows that the obtained mean difference is more than would be expected if there is no treatment effect. If it is large enough to be in the critical region, we conclude that the difference is **significant** or that the treatment has a significant effect. In this case we reject the null hypothesis. If the mean difference is relatively small, then the test statistic will have a low value. In this case, we conclude that the evidence from the sample is not sufficient, and the decision is fail to reject the null hypothesis.

ALPHA LEVEL



DIRECTIONAL TESTS

Includes the directional prediction in the statement of the hypotheses and in the location of the critical region.

For example, if the original population has a mean of $\mu = 80$ and the treatment is predicted to increase the scores, then the null hypothesis would state that after treatment:

H0: $\mu < 80$ (there is no increase)

In this case, the entire critical region would be located in the right-hand tail of the distribution because large values for M would demonstrate that there is an increase and would tend to reject the null hypothesis.

POWER OF A HYPOTHESIS TEST

The **power** of a hypothesis test is defined is the probability that the test will reject the null hypothesis when the treatment does have an effect. The power of a test depends on a variety of factors including the size of the treatment effect and the size of the sample.