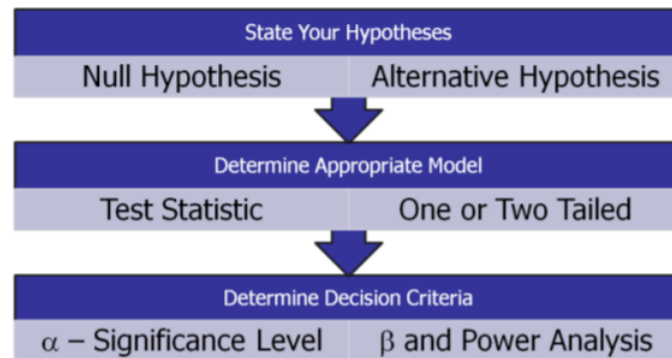


10.3: Design Research Hypotheses and Experiment

After developing a general question and having some sense of the data that is available or that is collected, we then design an experiment and a set of **hypotheses**.



Hypotheses and Hypothesis

Testing For purposes of testing, we need to design hypotheses that are statements about population parameters. Some examples of hypotheses:

At least 20% of juvenile offenders are caught and sentenced to prison.

- The mean monthly income for college graduates is over \$5000.
- The mean standardized test score for schools in Cupertino is the same as the mean scores for schools in Los Altos.
- The lung cancer rates in California are lower than the rates in Texas.
- The standard deviation of the New York Stock Exchange today is greater than 10 percentage points per year.

These same hypotheses could be written in symbolic notation:

- $p \geq 0.20$
- $\mu > 5000$
- $\mu_1 = \mu_2$
- $p_1 < p_2$
- $\sigma > 10$

Hypothesis Testing is a procedure, based on sample evidence and probability theory, used to determine whether the hypothesis is a reasonable statement and should not be rejected, or is unreasonable and should be rejected. This hypothesis that is tested is called the **Null Hypothesis** and is designated by the symbol H_0 . If the Null Hypothesis is unreasonable and needs to be rejected, then the research supports an **Alternative Hypothesis** designated by the symbol H_a .

Definition: Null Hypothesis (H_0)

A statement about the value of a population parameter that is assumed to be true for the purpose of testing.

Definition: Alternative Hypothesis (H_a)

A statement about the value of a population parameter that is assumed to be true if the Null Hypothesis is rejected during testing.

From these definitions it is clear that the Alternative Hypothesis will necessarily contradict the Null Hypothesis; both cannot be true at the same time. Some other important points about hypotheses:

- Hypotheses must be statements about population parameters, never about sample statistics.
- In most hypotheses tests, equality ($=$, \leq , \geq) will be associated with the Null Hypothesis while non-equality (\neq , $<$, $>$) will be associated with the Alternative Hypothesis.

- It is the Null Hypothesis that is always tested in attempt to “disprove” it and support the Alternative Hypothesis. This process is analogous in concept to a “proof by contradiction” in Mathematics or Logic, but supporting a hypothesis with a level of confidence is not the same as an absolute mathematical proof.

Examples of Null and Alternative Hypotheses:

- $H_o : p \leq 0.20$ $H_a : p > 0.20$
- $H_o : \mu \leq 5000$ $H_a : \mu > 5000$
- $H_o : \mu_1 = \mu_2$ $H_a : \mu_1 \neq \mu_2$
- $H_o : p_1 \geq p_2$ $H_a : p_1 < p_2$
- $H_o : \sigma \leq 10$ $H_a : \sigma > 10$

Statistical Model and Test Statistic

To test a hypothesis we need to use a **statistical model** that describes the behavior for data and the type of population parameter being tested. Because of the Central Limit Theorem, many statistical models are from the Normal Family, most importantly the Z , t , χ^2 , and F distributions. Other models that are used when the Central Limit Theorem is not appropriate are called non-parametric Models and will not be discussed here.

Each chosen model has requirements of the data called **model assumptions** that should be checked for appropriateness. For example, many models require that the sample mean have approximately a Normal Distribution, something that may not be true for some smaller or heavily skewed data sets.

Once the model is chosen, we can then determine a **test statistic**, a value derived from the data that is used to decide whether to **reject** or **fail to reject** the Null Hypothesis.

Examples of Statistical Models and Test Statistics

Statistical Model	Test Statistic
Mean vs. Hypothesized Value	$t = \frac{\bar{X} - \mu_o}{s/\sqrt{n}}$
Proportion vs. Hypothesized Value	$Z = \frac{\hat{p} - p_o}{\sqrt{\frac{p_o(1-p_o)}{n}}}$
Variance vs. Hypothesized Value	$\chi^2 = \frac{(n-1)s^2}{\sigma^2}$

Errors in Decision Making

Whenever we make a decision or support a position, there is always a chance we make the wrong choice. The hypothesis testing process requires us to either to reject the Null Hypothesis and support the Alternative Hypothesis or fail to reject the Null Hypothesis. This creates the possibility of two types of error:

- Type I Error Rejecting the null hypothesis when it is actually true.
- Type II Error Failing to reject the null hypothesis when it is actually false.

	Fail to Reject H_o	Reject H_o
H_o is true	Correct Decision	Type I error
H_o is False	Type II error	Correct Decision

In designing hypothesis tests, we need to carefully consider the probability of making either one of these errors.

Example: Pharmaceutical research

Recall the two news stories discussed earlier. In the first story, a drug company marketed a suppository that was later found to be ineffective (and often dangerous) in treatment. Before marketing the drug, the company determined that the drug was effective in treatment, meaning that the company rejected a Null Hypothesis that the suppository had no effect on the disease. This is an example of Type I error.

In the second story, research was abandoned when the testing showed Interferon was ineffective in treating a lung disease. The company in this case failed to reject a Null Hypothesis that the drug was ineffective. What if the drug really was effective? Did the company make Type II error? Possibly, but since the drug was never marketed, we have no way of knowing the truth.

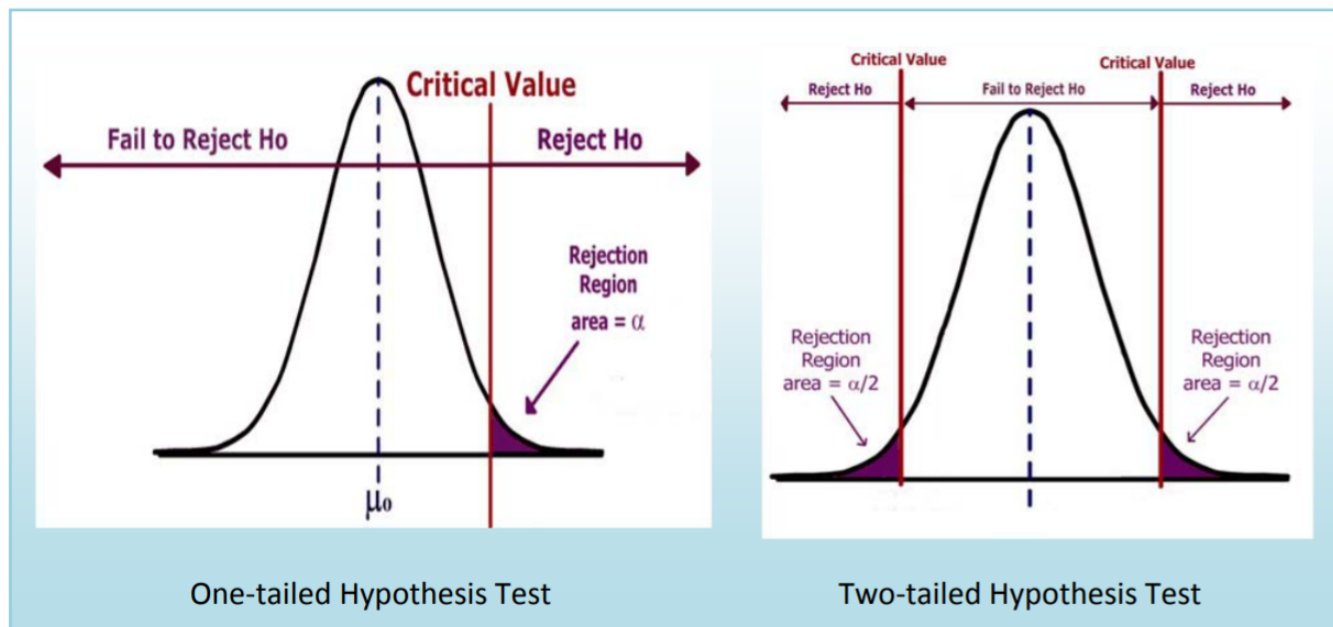
These stories highlight the problem of statistical research: errors can be analyzed using probability models, but there is often no way of identifying specific errors. For example, there are unknown innocent people in prison right now because a jury made Type I error in wrongfully convicting defendants. We must be open to the possibility of modification or rejection of currently accepted theories when new data is discovered.

In designing an experiment, we set a maximum probability of making Type I error. This probability is called the **level of significance** or **significance level** of the test and is designated by the Greek letter α , read as alpha. The analysis of Type II error is more problematic since there are many possible values that would satisfy the Alternative Hypothesis. For a specific value of the Alternative Hypothesis, the design probability of making Type II error is called **Beta (β)** which will be analyzed in detail later in this section.

Critical Value and Rejection Region

Once the significance level of the test is chosen, it is then possible to find the region(s) of the probability distribution function of the test statistic that would allow the Null Hypothesis to be rejected. This is called the **Rejection Region**, and the boundary between the Rejection Region and the “Fail to Reject” is called the **Critical Value**.

There can be more than one critical value and rejection region. What matters is that the total area of the rejection region equals the significance level α .



One and Two tailed Tests

A test is one-tailed when the Alternative Hypothesis, H_a , states a direction, such as:

H_o : The mean income of females is less than or equal to the mean income of males.

H_a : The mean income of females is greater than that of males.

Since equality is usually part of the Null Hypothesis, it is the Alternative Hypothesis which determines which tail to test.

A test is two-tailed when no direction is specified in the alternate hypothesis H_a , such as:

H_o : The mean income of females is equal to the mean income of males.

H_a : The mean income of females is not equal to the mean income of the males.

In a two tailed-test, the significance level is split into two parts since there are two rejection regions. In hypothesis testing, in which the statistical model is symmetrical (eg: the Standard Normal Z or Student's t distribution) these two regions would be equal. There is a relationship between a confidence interval and a two-tailed test: if the level of confidence for a confidence interval is equal to $1 - \alpha$, where α is the significance level of the two-tailed test, the critical values would be the same.

Here are some examples for testing the mean μ against a hypothesized value μ_0 :

Note

$H_a : \mu > \mu_0$ means test the upper tail and is also called a right-tailed test.

$H_a : \mu < \mu_0$ means test the lower tail and is also called a left-tailed test.

$H_a : \mu \neq \mu_0$ means test both tails.

Deciding when to conduct a one or two-tailed test is often controversial and many authorities even go so far as to say that only two-tailed tests should be conducted. Ultimately, the decision depends on the wording of the problem. If we want to show that a new diet reduces weight, we would conduct a lower tailed test, since we don't care if the diet causes weight gain. If instead, we wanted to determine if mean crime rate in California was different from the mean crime rate in the United States, we would run a two-tailed test, since different implies greater than or less than.

This page titled [10.3: Design Research Hypotheses and Experiment](#) is shared under a [CC BY-SA 4.0](#) license and was authored, remixed, and/or curated by [Maurice A. Geraghty](#) via [source content](#) that was edited to the style and standards of the LibreTexts platform.