* 1. **Significance Tests: The Basics**

Confidence intervals are one of the two most common types of statistical inference. Use a confidence interval when your goal is to estimate a population parameter. The second common type of inference, called significance tests, has a different goal: to assess the evidence provided by data about some claim concerning a population.

A **significance test** is a formal procedure for comparing observed data with a claim (also called a hypothesis) whose truth we want to assess. The claim is a statement about a parameter, like the population proportion p or the population mean µ. We express the results of a significance test in terms of a probability that measures how well the data and the claim agree.

**Example**:

Suppose a basketball player claimed to be an 80% free-throw shooter. To test this claim, we have him attempt 50 free-throws. He makes 32 of them. His sample proportion of made shots is 32/50 = 0.64. What can we conclude about the claim based on this sample data?

1. What is the population parameter we want to test?
2. What hypothesis do we want to test (in symbols and words)?
3. What evidence do we have (assume conditions of random independent and normal are met)?

We can use software to simulate 400 sets of 50 shots assuming that the player is really an 80% shooter.



You can say how strong the evidence against the player’s claim is by giving the probability that he would make as few as 32 out of 50 free throws if he really makes 80% in the long run.

1. Do we have enough evidence to reject our null hypothesis?

The observed statistic is so unlikely if the actual parameter value is p = 0.80 that it gives convincing evidence that the player’s claim is not true.

**The Reasoning of Significance Tests**

In reality, there are two possible explanations for the fact that he made only 64% of his free throws:

1. The player’s claim is correct (p = 0.8), and by bad luck, a very unlikely outcome occurred.
2. The population proportion is actually less than 0.8, so the sample result is not an unlikely outcome.

**An outcome that would rarely happen if a claim were true is good evidence that the claim is not true**

**Stating Hypotheses**

The claim tested by a statistical test is called the **null hypothesis** **(Ho)**. The test is designed to assess the strength of the evidence against the null hypothesis. Often the null hypothesis is a statement of “no difference.”

The claim about the population that we are trying to find evidence for is the **alternative hypothesis (Ha).**

In any significance test, the null hypothesis has the form

**Ho : parameter = value**

The alternative hypothesis has one of the forms

**Ha : parameter < value**

**Ha : parameter > value**

**Ha : parameter ≠ value**

To determine the correct form of Ha, read the problem carefully.

Hypotheses always refer to a population, not to a sample. Be sure to state Ho and Ha in terms of population parameters.

 It is never correct to write a hypothesis about a sample statistic, such as 

Example : Job Satisfaction

Does the job satisfaction of assembly-line workers differ when their work is machine-paced rather than self-paced? The response variable is the difference in satisfaction scores, self-paced minus machine-paced. Describe the parameter of interest and the hypotheses.

Example : Defective parts

A company claims that only 1% of their USB cables are defective. A consumer group believes that the defective rate is higher. Describe the parameter of interest and the hypotheses.

**Interpreting P-Values**

The probability, computed assuming H0 is true, that the statistic would take a value as extreme as or more extreme than the one actually observed is called the **P-value** of the test. The smaller the P-value, the stronger the evidence against H0 provided by the data.

* Small P-values are evidence against Ho because they say that the observed result is unlikely to occur when Ho is true.
* Large P-values fail to give convincing evidence against Ho because they say that the observed result is likely to occur by chance when Ho is true.

Example : Job Satisfaction

Does the job satisfaction of assembly-line workers differ when their work is machine-paced rather than self-paced? The response variable is the difference in satisfaction scores, self-paced minus machine-paced. Describe the parameter of interest and the hypotheses.

For the job satisfaction study, the hypotheses are

Ho: µ = 0

Ha: µ ≠ 0



1. Explain what it means for the null hypothesis to be true in this setting.
2. Interpret the P-value in context.

**Statistical Significance**

The final step in performing a significance test is to draw a conclusion about the competing claims you were testing.

We will make one of two decisions based on the strength of the evidence against the null hypothesis (and in favor of the alternative hypothesis) -- reject H0 or fail to reject H0.

* If our sample result is too unlikely to have happened by chance assuming H0 is true, then we’ll reject H0.
* Otherwise, we will fail to reject H0.

A fail-to-reject H0 decision in a significance test doesn’t mean that H0 is true. For that reason, you should never “accept H0” or use language implying that you believe H0 is true.

There is no rule for how small a P-value we should require in order to reject H0 — it’s a matter of judgment and depends on the specific circumstances. But we can compare the P-value with a fixed value that we regard as decisive, called the **significance level**. We write it as α, the Greek letter alpha. When our P-value is less than the chosen α, we say that the result is **statistically significant**.

If the P-value is smaller than alpha, we say that the data are statistically significant at level α. In that case, we reject the null hypothesis H0 and conclude that there is convincing evidence in favor of the alternative hypothesis Ha.

**When we use a fixed level of significance to draw a conclusion in a significance test,**

**P-value < α → reject H**0 **→ conclude H**a **(in context)**

**P-value ≥ α → fail to reject H**0 **→ cannot conclude H**a **(in context)**

Example: Better Batteries

A company has developed a new deluxe AAA battery that is supposed to last longer than its regular AAA battery. However, these new batteries are more expensive to produce, so the company would like to be convinced that they really do last longer. Based on years of experience, the company knows that its regular AAA batteries last for 30 hours of continuous use, on average. The company selects an SRS of 15 new batteries and uses them continuously until they are completely drained. A significance test is performed using the hypotheses

H0 : µ = 30 hours

Ha : µ > 30 hours

where µ is the true mean lifetime of the new deluxe AAA batteries. The resulting P-value is 0.0276.

1. What conclusion can you make for the significance level α = 0.05?
2. What conclusion can you make for the significance level α = 0.01?