Lesson 10-2 Tests of Significance

# Significance Tests

A **test of significance** is intended to assess the evidence provided by data against a **null hypothesis** Ho in favor of an **alternative hypothesis** Ha.

The hypotheses are stated in terms of population parameters. Usually Ho is a statement that no effect is present, and Ha says that a parameter differs from its null value in a specific direction (**one-sided alternative**) or in either direction (**two-sided alternative**).

# The essential reasoning of a hypothesis test goes as follows:

1. Describe the effect you are searching for in terms of a population parameter like the mean µ. Never state the hypothesis in terms of a sample statistic like Ë.
2. The null hypothesis is the statement that this effect is not present in the population.
3. From the data, calculate a statistic like Ë **(test statistic)** that estimates the parameter. Is the value of this statistic far from the parameter value stated by the null hypothesis? If so, the data give evidence that the null hypothesis is false and that the effect that you are looking for is really there.
4. The **P-value** says how unlikely a result at least as extreme as the one we observed would be if the null hypothesis is true. Results with small P-values would rarely occur if the null hypothesis were true. We call such results **statistically significant.**

# Use the Steps for a test of significance

1. State the hypotheses in the context of the problem.
2. Calculate the test statistic.
3. State which method you will be using, check for conditions, and find the P-value.
4. State the conclusion *in the context of the problem*.

Before we begin to a test of significance, we often announce in advance what level of evidence we will insist upon. The decisive value of P is called the **significance level**. We write it as α. If, for example, if we choose α = 0.05, we are requiring that the data give evidence against Ho so strong that it would happen no more than 5% of the time (1 time in 20) when Ho is true. If we choose α = 0.01, we are insisting on stronger evidence against H0, evidence so strong that it would appear only 1% of the time (1 time in 100) if Ho is in fact true.

# What Can Go Wrong?

* Statistically significant does not mean practically significant.
* Even very small differences in P-values can be important if there is a very large sample involved.
* Significance tests are not always valid. Faulty data collection, outliers in the data, and not correctly replicating the experiment can invalidate the test.