

Hypothesis Testing

This lesson will focus on the basics of hypothesis testing and how to perform different types of hypothesis tests.

WE'LL COVER THE FOLLOWING



- Hypothesis testing
 - Null and Alternate hypothesis
 - The testing process

Hypothesis testing

During our analysis of the different datasets, we are often concerned with questions like whether males default more than females? Do self-driving cars crash more than normal cars? Does drug X help prevent/treat disease Y? To answer these questions, we can use another statistical technique known as **Hypothesis Testing**.

During data exploration, we discovered interesting patterns hidden in the data. Hypothesis testing enables us to confirm whether these patterns were present in the data by luck or by some real phenomena.

Null and Alternate hypothesis

The aim of the hypothesis test is to determine whether the null hypothesis can be rejected or not. The **null hypothesis** is a statement that assumes that nothing interesting is going on, or no relationship is present between two variables, or that there is no difference between a sample and a population.

For instance, if we suspect that males default more than females, the null hypothesis would be that males do not default more than females. If there is little or no evidence against the null hypothesis, we accept the null hypothesis. Otherwise, we reject the null hypothesis in favor of the **alternate hypothesis**, which states that something interesting is going on, or there is a relationship between two variables, or that there is a difference between a sample and a population.

between two variables, or that the sample is different from the population.

To reiterate, the null hypothesis is assumed true and statistical evidence is required to reject it in favor of the alternative hypothesis.

The testing process

The testing process is:

- State both hypotheses.
- Choose the **significance level (α)**. It is the probability threshold that determines whether you accept or reject a hypothesis.
- Choose a test and compute the test statistic.
- Make the decision to reject or accept the null hypothesis based on if the probability of getting a result as extreme as the one you observe. If it is **lower** than the significance level, you reject the null hypothesis in favor of the alternative. This probability of seeing a result as extreme or more extreme than the one observed is known as the **p-value**.

Usually, three types of tests are carried out:

- One-Sample t-Test
- Two-Sample t-Test
- Paired t-Test

We will look at these in the next few lessons starting from *one-sample t-test* in the next lesson.