HYPOTESIS TESTING

Hypothesis testing is a technique that helps scientists, researchers, or for that matter, anyone test the validity of their claims or hypotheses about **real-world** or **real-life** events. Hypothesis testing techniques are often used in statistics and **data science** to analyze whether the claims about the occurrence of the events are true, whether the results returned by performance metrics of machine learning models are representative of the models or they happened by chance. This blog post will cover some of the key statistical concepts including steps and examples in relation to what is **hypothesis testing**, and, how to **formulate them**. The knowledge of hypothesis formulation and hypothesis testing holds the key to establish truth about the hypotheses and modify business processes for maximum business impact.

What is Hypothesis Testing?

**The hypothesis can be defined as the claim that can either be related to the truth about something that exists in the world, or, truth about something that’s needs to be established a fresh**. In simple words, another word for the hypothesis is the “**claim**”. Until the claim is proven to be true, it is called the hypothesis. Once the claim is proved, it becomes the new truth about the thing. For example, let’s say that a claim is made that students studying for more than 6 hours a day gets more than 90% of marks in their examination. Now, this is just a claim or a hypothesis and not the truth in the real world. However, in order for the claim to become the truth for widespread adoption, it needs to be proved using pieces of evidence, e.g., data. In order to reject this claim or otherwise, one needs to do some empirical analysis by gathering data samples and evaluating the claim.

**The process of gathering data and evaluating the claims or hypotheses with the goal to reject or otherwise (failing to reject) is termed hypothesis testing**. Note the wordings – “**failing to reject**”. **It means that we don’t have enough evidence to reject the claim**. Thus, until the time that new evidence comes up, the claim can be considered the truth. There are different techniques to test the hypothesis in order to reach the conclusion of whether the hypothesis can be used to represent the truth of the world. One must note that the hypothesis testing never constitutes a proof that the hypothesis is absolute truth based on the observations. It only provides added support to consider the hypothesis as truth until the time that new evidences can against the hypotheses can be gathered. We can never be 100% sure about truth related to those hypotheses based on the hypothesis testing.

Simply speaking, hypothesis testing is a framework that can be used to assert whether the claim or the hypothesis made about a real-world/real-life event can be seen as the truth or otherwise based on the given data (evidences). For example:

* It is claimed that a 500 gm sugar packet for a particular brand, say XYZA, contains sugar of less than 500 gm, say around 480gm.  Can this claim be taken as truth? How do we know that this claim is true? This is a hypothesis until proved.
* A group of doctors claims that quitting smoking increases lifespan. Can this claim be taken as new truth? The hypothesis is that quitting smoking results in an increase in lifespan
* It is claimed that brisk walking for half an hour every day reverses diabetes. In order to accept this in your lifestyle, you may need evidence that supports this claim or hypothesis.
* One common real-life example of hypothesis testing is **election polling**. In order to predict the outcome of an election, pollsters take a sample of the population and ask them who they plan to vote for. They then use hypothesis testing to assess whether their sample is **representative of the population as a whole**. **If the results of the hypothesis test are significant, it means that the sample is representative and that the poll can be used to predict the outcome of the election**. However, if the results are not significant, it means that the sample is not representative and that the poll should not be used to make predictions.

Now that the hypothesis is stated, let’s go ahead and state and formulate the hypothesis as the null and alternate hypothesis in order to perform hypothesis testing.

HYPOTESIS TESTING EXAMPLE

Before we get ahead and start understanding more details about hypothesis and hypothesis testing steps, lets take a look at some real-world examples of how to think about hypothesis and hypothesis testing when dealing with real-world problems:

**Customer churn**: Customer churn is one of the most common problem one come across when starting to work with AI / machine learning. Customer churn refers to the process of customers leaving a company or service. It is the percentage of customers who stop doing business with a company in a given time period. Not only does it lose revenue from the customer who leaves, but it also incurs the cost of acquiring a new customer. Thus, business wants to take action to stop the customer churn. In order to take one or more actions, they need to make decisions which will be followed by those actions. These decisions are based on claims or hypothesis made by the business stakeholders in relation to why customer are leaving the services. This is where hypothesis and hypothesis testing comes into picture. Let’s look at some of the hypothesis which can be put to hypothesis testing and later carved into analytical solutions such as dashboards, or AI / machine learning solutions as appropriate.

* Customers are churning because they ain’t getting response to their complaints or issues
* Customers are churning because there are other competitive services in the market which are providing these services at lower cost.
* Customers are churning because there are other competitive services which are providing more services at the same cost

You may note different hypotheses which are listed above. The next step would be validate some of these hypotheses. This is where data scientists will come into picture. One or more data scientists may be asked to work on different hypotheses. This would result in these data scientists looking for appropriate data related to the hypothesis they are working. This section will be detailed out in near future.

STATE THE HYPOTHESIS TO BEGIN HYPOTHESIS TESTING

**The first step to hypothesis testing is defining or stating a hypothesis.** Before the hypothesis can be tested, we need to formulate the hypothesis in terms of mathematical expressions. There are two important aspects to pay attention to, prior to the formulation of the hypothesis. The following represents different types of hypothesis that could be put to hypothesis testing:

* **Claim made against the well-established fact**: The case in which a fact is well-established, or accepted as truth or “knowledge” and a new claim is made about this well-established fact. For **example**, when you buy a packet of 500 gm of sugar, you assume that the packet does contain at the minimum 500 gm of sugar and not any less, based on the label of 500 gm on the packet. In this case, the fact is given or assumed to be the truth. A new claim can be made that the 500 gm sugar contains sugar weighing less than 500 gm. This claim needs to be tested before it is accepted as truth. Such cases could be considered for hypothesis testing if this is claimed that the assumption or the default state of being is not true.
* **Claim to establish the new truth**: The case in which there is some claim made about the reality that exists in the world (fact). For **example**, the fact that the housing price depends upon the average income of people already staying in the locality can be considered as a claim and not assumed to be true. Another **example** could be the claim that running 5 miles a day would result in a reduction of 10 kg of weight within a month. There could be varied such claims which when required to be proved as true have to go through hypothesis testing.

Based on the above considerations, the following hypothesis can be stated for doing hypothesis testing.

* The packet of 500 gm of sugar contains sugar of weight less than 500 gm. (**Claim made against the established fact**)
* The housing price depends upon the average income of the people staying in the locality. (**Claim to establish new truth**)
* Running 5 miles a day results in a reduction of 10 kg of weight within a month. (**Claim to establish new truth**)

FORMULATE NULL & ALTERNATE HYPOTHESIS AS NEXT STEP

Once the hypothesis is defined or stated, the next step is to **formulate** the null and **alternate hypothesis** in order to begin hypothesis testing as described above.

1. What is Null Hypotesis?

In the case where the given statement is a well-established fact or default state of being in the real world, one can call it a **null hypothesis** (in the simpler word, nothing new). Well-established facts don’t need any hypothesis testing and hence can be called the **null hypothesis.**In cases, when there are any new claims made which is not well established in the real world, the null hypothesis can be thought of as the default state or opposite state of that claim. For **example**, in the previous section, the claim or hypothesis is made that the students studying for more than 6 hours a day gets more than 90% of marks in their examination. The null hypothesis, in this case, will be that the claim is not true or real. The null hypothesis can be stated as the fact that it is not a truth that the students reading more than 6 hours a day would get more than 90% of the marks. Another **example** of **hypothesis** is when somebody is alleged that they have performed a crime. The default state of the world is that the person has not committed the crime and he/she is guilty. This will be null hypothesis.

1. What is an alternate hypothesis?

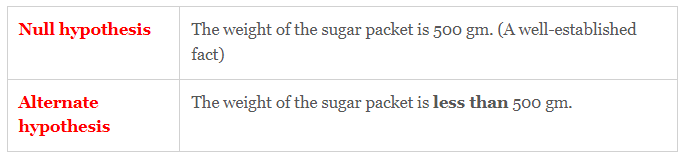
In case the given statement is a claim (unexpected event in the real world) and not yet proven, one can call/formulate it as an **alternate hypothesis** and accordingly define a null hypothesis which is the opposite state of the hypothesis. In simple words, the hypothesis or claim that needs to be tested against reality in the real world can be termed the alternate hypothesis. In order to reach a conclusion that the claim (alternate hypothesis) can be considered the new truth (based on the available evidence), it would be important to reject the null hypothesis. It should be noted that null and alternate hypotheses are mutually exclusive and at the same time asymmetric. In the **example** given in the previous section, the claim that the students studying for more than 6 hours get more than 90% of marks can be termed as the alternate hypothesis.

Once the hypothesis is formulated as null and alternate hypothesis, there are two possible outcomes that can happen from hypothesis testing as a function of null and alternate hypothesis. These outcomes are the following:

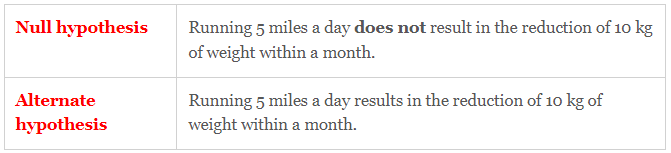
* **Reject the null hypothesis**: There is **enough evidence** based on which one can reject the null hypothesis. Let’s understand this with the help of an **example** provided earlier in this section. The null hypothesis is that there is no relationship between the students studying more than 6 hours a day and getting more than 90% marks. In a sample of 30 students studying more than 6 hours a day, it was found that they scored 91% marks. Given that the null hypothesis is true, this kind of hypothesis testing result will be highly unlikely. This kind of result can’t happen by chance. That would mean that the claim can be taken as the new truth in the real world. One can go and take further samples of 30 students to perform some more testing to validate the hypothesis. If similar results show up with other tests, it can be said with very high confidence that there is enough evidence to reject the null hypothesis that there is no relationship between the students studying more than 6 hours a day and getting more than 90% marks. In such cases, one can go to accept the claim as new truth that the students studying more than 6 hours a day get more than 90% marks. The hypothesis can be considered the new truth until the time that new tests provide evidence against this claim.
* **Fail to reject the null hypothesis**: There is **not enough evidence**-based on which one can reject the null hypothesis (well-established fact or reality). Thus, one would fail to reject the null hypothesis. In a sample of 30 students studying more than 6 hours a day, the students were found to score 75%. Given that the null hypothesis is true, this kind of result is fairly likely or expected. With the given sample, one can’t reject the null hypothesis that there is no relationship between the students studying more than 6 hours a day and getting more than 90% marks.

1. Examples of formulating the null and alternate hypothesis

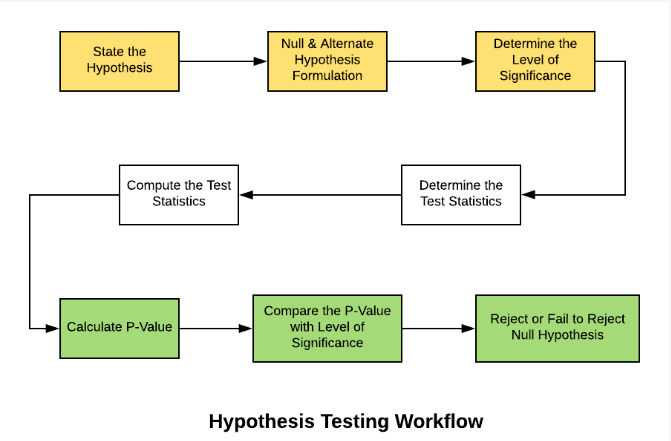
Take the **example** of sugar with the label 500 gm. As per the above, this represents the scenario when **the statement made is believed to be true in reality**. Thus, it is believe to be true (based on the given label) that the sugar packet weighs 500 gm. The claim, however, is made that the sugar packet with label as 500gm weighs significantly less than 500 gm (around 480 gm). Thus, we will need to do hypothesis testing to find out whether the claim made is true or otherwise. The hypothesis testing would need to be done to ascertain the truth about the label mentioned as 500 gm because there is a claim that sugar packets consisted of 480 gm. In this scenario, the **null hypothesis** would get formulated as the statement that the weight of canned sugar is equal to 500 gm. The alternate hypothesis will thus get formulated as the statement that the weight of the sugar packet is less than 500 gm.



Take the **example** of a claim that running 5 miles a day will lead to a reduction of 10 kg of weight within a month. Now, **this is the hypothesis or claim which is required to be proved or otherwise**. The **alternate hypothesis** will be formulated first as the statement that “running 5 miles a day will lead to a reduction of 10 kg of weight within a month”. Hence, the null hypothesis will be the opposite of the alternate hypothesis and stated as the fact that “running 5 miles a day does not lead to a reduction of 10 kg of weight within a month”.



Hypothesis Testing Steps



Based on the above, the following are some of the**steps** to be taken when doing hypothesis testing:

* **State the hypothesis**: First and foremost, the hypothesis needs to be stated. The hypothesis could either be the statement that is assumed to be true or the claim which is made to be true.
* **Formulate the hypothesis**: This step requires one to identify the Null and Alternate hypotheses or in simple words, formulate the hypothesis. Take an **example** of the canned sauce weighing 500 gm as the Null Hypothesis.
* **Set the criteria for a decision**: Identify test statistics that could be used to assess the Null Hypothesis. The test statistics with the above **example** would be the average weight of the sugar packet, and **t-statistics** would be used to determine the **P-value**. For different kinds of problems, different kinds of statistics including **Z-statistics, T-statistics, F-statistics**, etc can be used.
* **Identify the level of significance (alpha)**: Before starting the hypothesis testing, one would be required to set the significance level (also called as *alpha*) which represents the value for which a P-value less than or equal to *alpha* is considered statistically significant. Typical values of *alpha* are 0.1, 0.05, and 0.01. **In case the P-value is evaluated as statistically significant, the null hypothesis is rejected. In case, the P-value is more than the *alpha* value, the null hypothesis is failed to be rejected.**
* **Compute the test statistics**: Next step is to calculate the test statistics (z-test, t-test, f-test, etc) to determine the P-value. If the sample size is more than 30, it is recommended to use z-statistics. Otherwise, t-statistics could be used. In the current **example** where 20 packets of canned sauce is selected for hypothesis testing, t-statistics will be calculated for the mean value of 505 gm (sample mean). The t-statistics would then be calculated as the difference of 505 gm (sample mean) and the population means (500 gm) divided by the sample standard deviation divided by the square root of sample size (20).
* **Calculate the P-value of the test statistics**: Once the test statistics have been calculated, find the P-value using either of t-table or a z-table. P-value is the probability of obtaining a test statistic (t-score or z-score) equal to or more extreme than the result obtained from the sample data, given that the null hypothesis H0 is true.
* **Compare P-value with the level of significance (Alpha/0,05)**: The significance level is set as the allowable range within which if the value appears, one will be **failed to reject** the Null Hypothesis. This region is also called as **Non-rejection region**. The value of alpha is compared with the p-value. **If the p-value is less than the significance level, the test is statistically significant and hence, the null hypothesis will be rejected.**

P-Value: Key to Statistical Hypothesis Testing

Once you formulate the hypotheses, there is the need to test those hypotheses. Meaning, say that the null hypothesis is stated as the statement that housing price does not depend upon the average income of people staying in the locality, it would be required to be tested by taking samples of housing prices and, based on the test results, this Null hypothesis could either be **rejected** or **failed to be rejected**. In hypothesis testing, the following two are the outcomes:

* **Reject** the Null hypothesis
* **Fail to Reject** the Null hypothesis

Take the above **example** of the sugar packet weighing 500 gm. The Null hypothesis is set as the statement that the sugar packet weighs 500 gm. After taking a sample of 20 sugar packets and testing/taking its weight, it was found that the average weight of the sugar packets came to 495 gm. The test statistics (t-statistics) were calculated for this sample and the P-value was determined. Let’s say the P-value was found to be 15%. Assuming that the level of significance is selected to be 5%, the test statistic is not statistically significant (P-value > 5%) and thus, the null hypothesis fails to get rejected. Thus, one could safely conclude that the sugar packet does weigh 500 gm. However, if the average weight of canned sauce would have found to be 465 gm, this is way beyond/away from the mean value of 500 gm and one could have ended up **rejecting** the **Null Hypothesis based on the P-value**.