What is Hypothesis Testing and When Do We Use It?

Hypothesis testing is a statistical method used to evaluate a claim or hypothesis about a population parameter based on sample data. It involves making decisions about the validity of a statement, often referred to as the null hypothesis, by assessing the likelihood of observing the sample data if the null hypothesis were true.

This process helps researchers determine whether there is enough evidence to support or reject the null hypothesis, thereby drawing conclusions about the population of interest. In essence, hypothesis testing provides a structured approach for making inferences and decisions in the face of uncertainty, playing a crucial role in scientific research, data analysis, and decision-making across various domains.

Hypothesis testing is a part of statistical analysis and machine learning, where we test the assumptions made regarding a population parameter.

We use hypothesis testing in various scenarios, including:

1. **Scientific research:** Testing the effectiveness of a new drug, evaluating the impact of a treatment on patient outcomes, or examining the relationship between variables in a study.
2. **Quality control:**Assessing whether a manufacturing process meets specified standards or determining if a product’s performance meets expectations.
3. **Business decision-making:** Investigating the effectiveness of marketing strategies, analyzing customer preferences, or testing hypotheses about financial performance.
4. **Social sciences:** Studying the effects of interventions on societal outcomes, examining attitudes and behaviors, or testing theories about human behavior.

Note: Don’t be confused between the terms Parameter and Satistic.  
**A Parameter is a number that describes the data from the *population* whereas, a *Statistic* is a number that describes the data from a *sample*.**

In hypothesis testing, several key terms and concepts are commonly used to describe the process and interpret results:

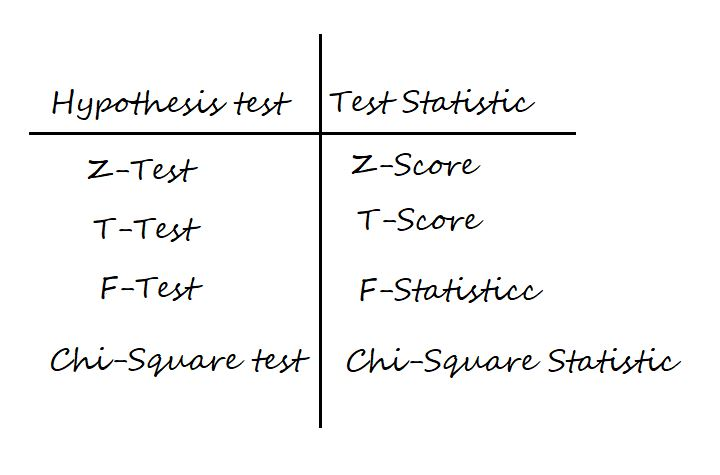
**1. Null Hypothesis (H0)**: Null hypothesis is a statistical theory that suggests there is no statistical significance exists between the populations. It is denoted by **H0** and read as **H-naught**.

**2. Alternative Hypothesis (Ha or H1):** An Alternative hypothesis suggests there is a significant difference between the population parameters. It could be greater or smaller. Basically, it is the contrast of the Null Hypothesis. It is denoted by **Ha** or **H1**.

**Note: H0 must always contain equality(=). Ha always contains difference(**≠, **>, <).**

For example, if we were to test the equality of average means (µ) of two groups:  
for a two-tailed test, we define H0: µ1 = µ2 and Ha: µ1≠µ2  
for a one-tailed test, we define H0: µ1 = µ2 and Ha: µ1 > µ2 or Ha: µ1 < µ2

**3. Test Statistic:**It is denoted by t and is dependent on the test that we run. It is the deciding factor to reject or accept theNull Hypothesis. The four main test statistics are given in the below table:



**4. Significance Level (α):** The significance level, often denoted by α (alpha), represents the probability of rejecting the null hypothesis when it is actually true. Commonly used significance levels include 0.05 and 0.01, indicating a 5% and 1% chance of Type I error, respectively.

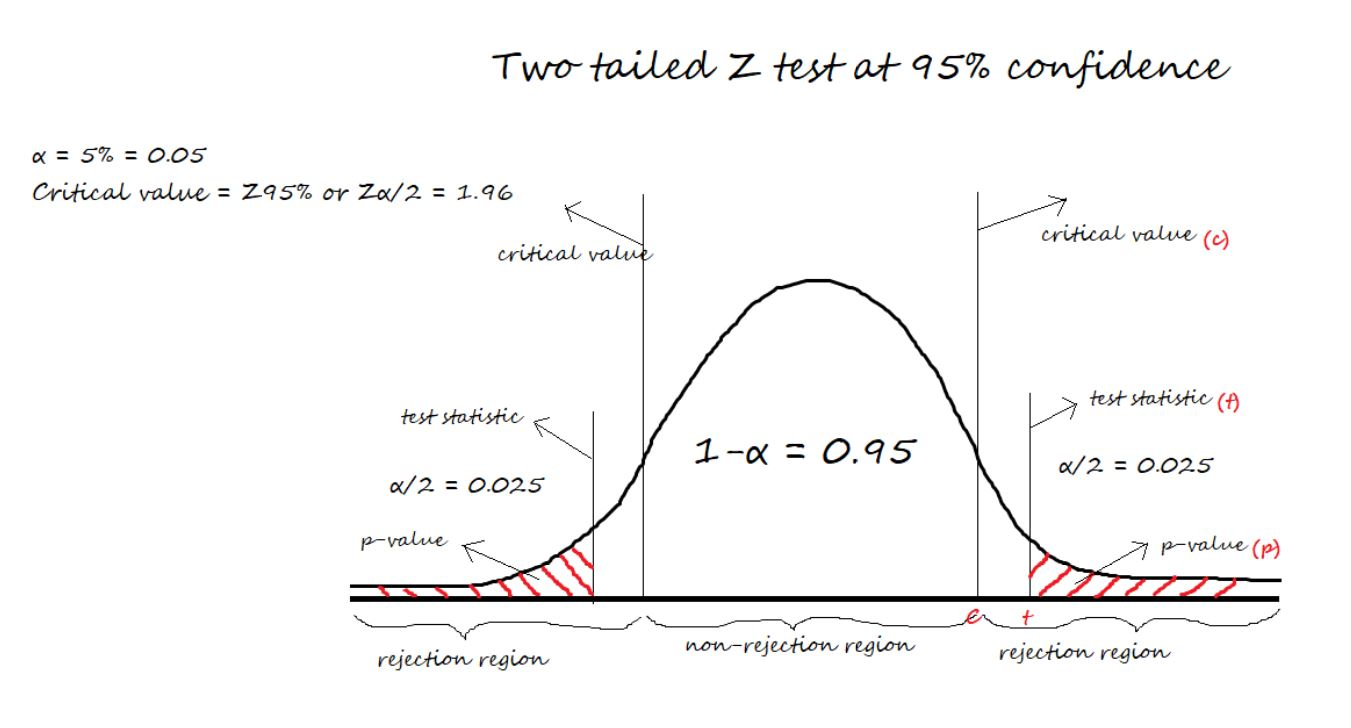
**5. P-value:** It is the proportion of samples (assuming the Null Hypothesis is true) that would be as extreme as the test statistic. It is denoted by the letter **p**.

**6. Critical Value:**Denoted by **C** and it is a value in the distribution beyond which leads to the rejection of the Null Hypothesis. It is compared to the test statistic.

Now, assume we are running a two-tailed Z-Test at 95% confidence. Then, the level of significance (α) = 5% = 0.05. Thus, we will have (1-α) = 0.95 proportion of data at the center, and α = 0.05 proportion will be equally shared to the two tails. Each tail will have (α/2) = 0.025 proportion of data.

The critical value i.e., Z95% or Zα/2 = 1.96 is calculated from the [Z-scores table](https://www.math.arizona.edu/~rsims/ma464/standardnormaltable.pdf).

Now, take a look at the below figure for a better understanding of critical value, test-statistic, and p-value.



Steps of Hypothesis Testing

The steps of hypothesis testing typically involve the following process:

1. **Formulate Hypotheses**: State the null hypothesis and the alternative hypothesis.
2. **Choose Significance Level (α)**: Select a significance level (α), which determines the threshold for rejecting the null hypothesis. Commonly used significance levels include 0.05 and 0.01.
3. **Select Appropriate Test**: Choose a statistical test based on the research question, type of data, and assumptions. Common tests include t-tests, chi-square tests, ANOVA, correlation tests, and regression analysis, among others.
4. **Collect Data and Calculate Test Statistic**: Collect relevant sample data and calculate the appropriate test statistic based on the chosen statistical test.
5. **Determine Critical Region**: Define the critical region or rejection region based on the chosen significance level and the distribution of the test statistic.
6. **Calculate P-value**: Determine the probability of observing a test statistic as extreme as, or more extreme than, the one obtained from the sample data, assuming the null hypothesis is true. The p-value is compared to the significance level to make decisions about the null hypothesis.
7. **Make Decision**: If the p-value is less than or equal to the significance level (p ≤ α), reject the null hypothesis in favor of the alternative hypothesis. If the p-value is greater than the significance level (p > α), fail to reject the null hypothesis.
8. **Draw Conclusion**: Interpret the results based on the decision made in step 7. Provide implications of the findings in the context of the research question or problem.
9. **Check Assumptions and Validate Results**: Assess whether the assumptions of the chosen statistical test are met. Validate the results by considering the reliability of the data and the appropriateness of the statistical analysis.

By following these steps systematically, researchers can conduct hypothesis tests, evaluate the evidence, and draw valid conclusions from their analyses.

Decision Rules

The two methods of concluding the Hypothesis test are using the Test-statistic value and p-value.

In both methods, we start assuming the Null Hypothesis to be true, and then we reject the Null hypothesis if we find enough evidence.

**The decision rule for the Test-statistic method:**

if test-statistic (t) > critical Value (C), we reject Null Hypothesis.  
If test-statistic (t) ≤ critical value (C), we fail to reject Null Hypothesis.

**The decision rule for the p-value method:**

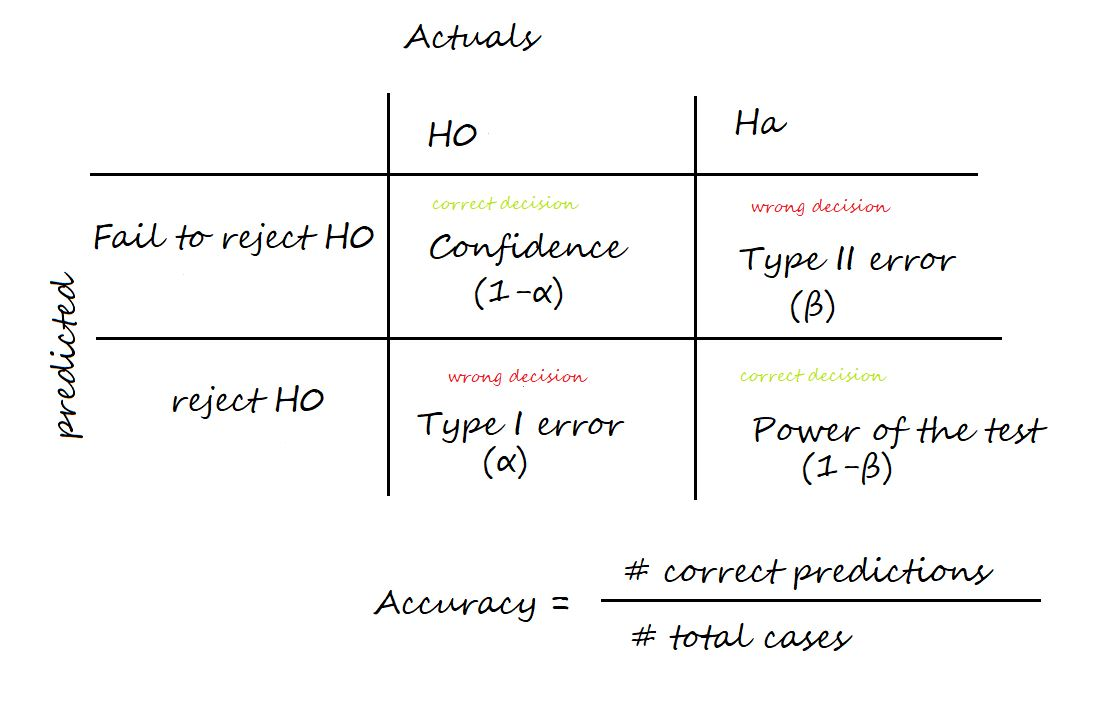
if p-value (p) > level of significance (α), we fail to reject Null Hypothesis  
if p-value (p) ≤ level of significance (α), we reject Null Hypothesis

 In easy terms, we say **P High, Null Fly,**and**P Low, Null Go**.

Confusion Matrix in Hypothesis Testing

To plot a confusion matrix, we can take actual values in columns and predicted values in rows or vice versa.

(I am illustrating by taking actuals in columns and predicting in rows.)



**Confidence:**The probability of accepting a True Null Hypothesis. It is denoted as (1-α)

**Power of test:**The probability of rejecting a False Null Hypothesis i.e., the ability of the test to detect a difference. It is denoted as (1-β) and its value lies between 0 and 1.

**Type I error:**Occurs when we reject a True Null Hypothesis and is denoted as α.

**Type II error:**Occurs when we accept a False Null Hypothesis and is denoted as β.

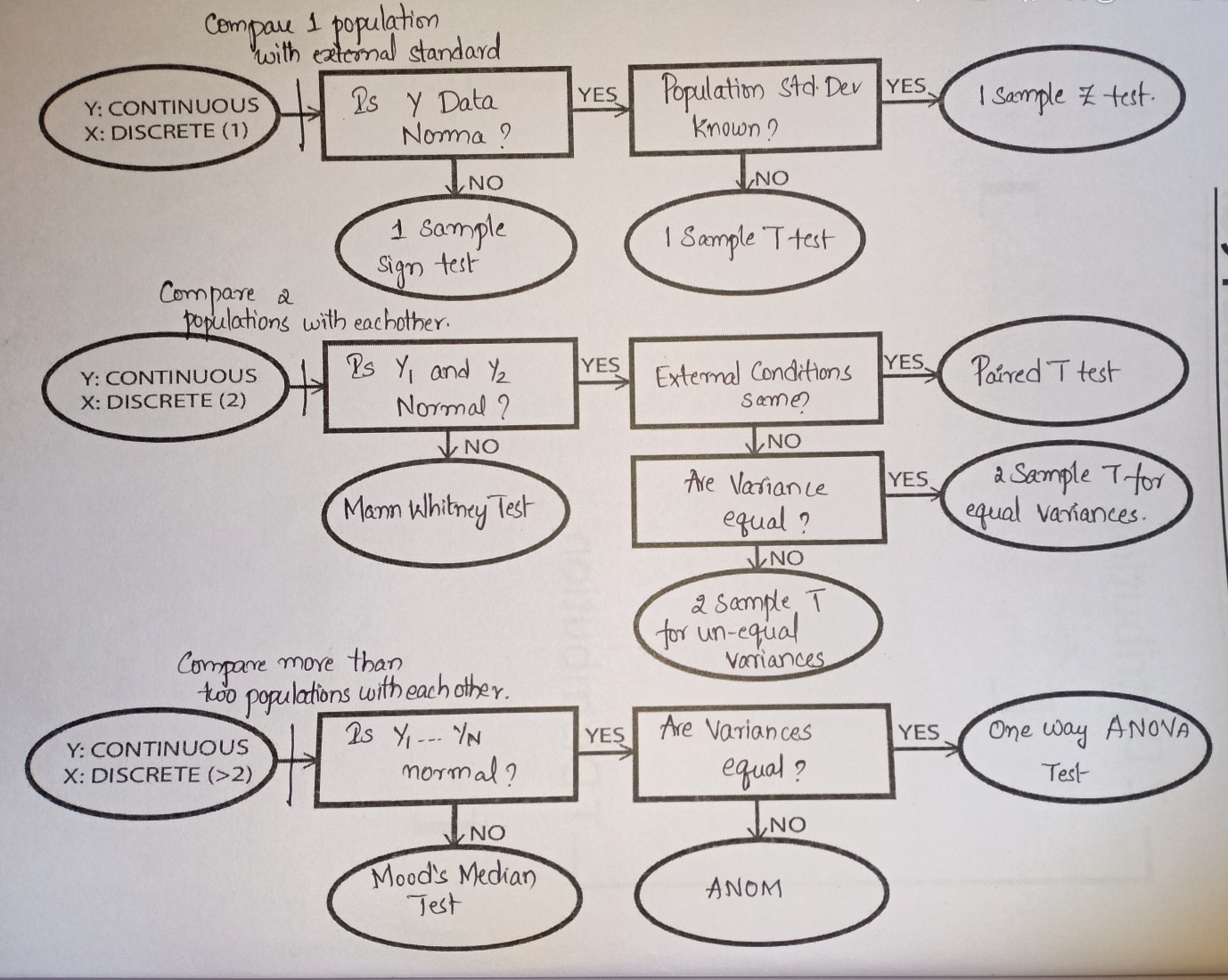
**Accuracy:** Number of correct predictions / Total number of cases

The factors that affect the power of the test are sample size, population variability, and the confidence (α).  
Confidence and power of test are directly proportional. Increasing the confidence increases the power of the test.

Types of Hypothesis Tests

In this section, we will see some examples of two different types of hypothesis tests.

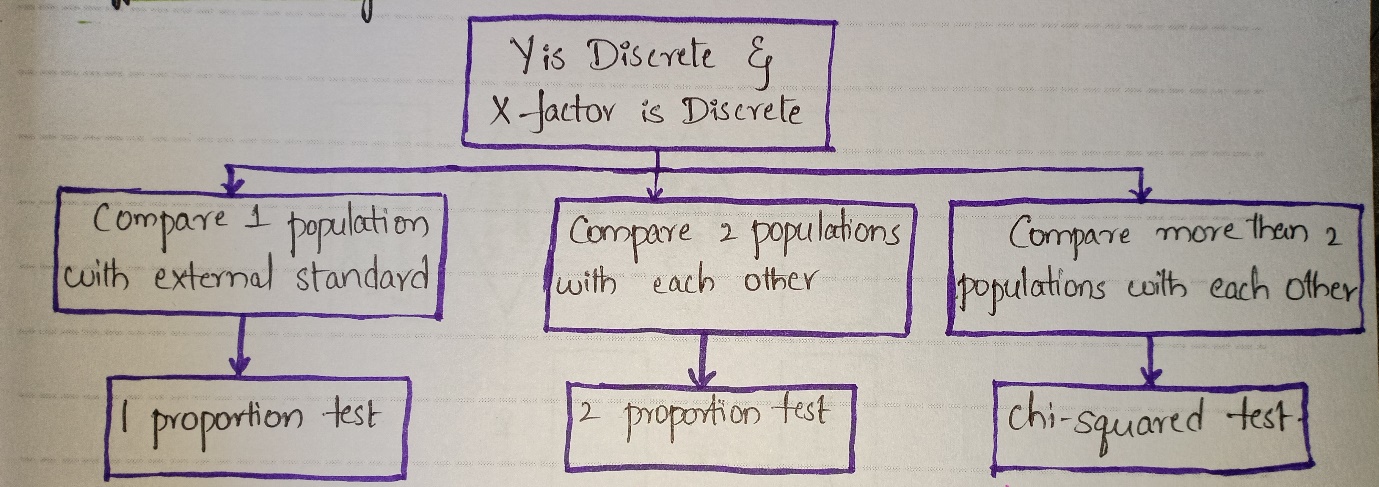
Hypothesis Tests When the Data is Continuous



When dealing with continuous data, several common hypothesis tests are used, depending on the research question and the characteristics of the data. Some of the most widely used hypothesis tests for continuous data include:

1. **One-Sample t-test**: Used to compare the mean of a single sample to a known value or hypothesized population mean.
2. **Paired t-test**: Compares the means of two related groups (e.g., before and after treatment) to determine if there is a significant difference.
3. **Independent Samples t-test**: Compares the means of two independent groups to determine if there is a significant difference between them.
4. **Analysis of Variance (ANOVA)**: Used to compare means across three or more independent groups to determine if there are any statistically significant differences.
5. **Correlation Test (Pearson’s correlation coefficient)**: Determines if there is a linear relationship between two continuous variables.
6. **Regression Analysis**: Evaluates the relationship between one dependent variable and one or more independent variables.

Hypothesis Tests When the Data is Discrete



When dealing with discrete data, several common hypothesis tests are used to analyze differences between groups, associations, or proportions. Some of the most widely used hypothesis tests for discrete data include:

1. **Chi-Square Test of Independence**: Determines whether there is a significant association between two categorical variables by comparing observed frequencies to expected frequencies.
2. **Chi-Square Goodness-of-Fit Test**: Assesses whether the observed frequency distribution of a single categorical variable differs significantly from a hypothesized or expected distribution.
3. **Binomial Test**: Determines whether the proportion of successes in a series of independent Bernoulli trials differs significantly from a hypothesized value.
4. **Poisson Test**: Tests whether the observed counts of events in a fixed interval of time or space follow a Poisson distribution, often used in count data analysis.
5. **McNemar’s Test**: Analyzes changes or differences in paired categorical data, typically used in before-and-after studies or matched case-control studies.
6. **Fisher’s Exact Test**: Determines the significance of the association between two categorical variables in small sample sizes when the assumptions of the chi-square test are not met.

These tests are valuable tools for analyzing categorical data, identifying relationships between variables, and making inferences about populations based on sample data. The choice of test depends on the research question, the nature of the data, and the study design.

Types of Errors in Hypothesis Testing

In hypothesis testing, there are two main types of errors:

1. **Type I error (False Positive):** This occurs when the null hypothesis is incorrectly rejected, indicating a significant result when there is actually no true effect or difference in the population being studied.
2. **Type II error (False Negative):** This occurs when the null hypothesis is incorrectly retained, failing to reject it when there is actually a true effect or difference in the population being studied.

These errors represent the trade-off between making incorrect conclusions and the risk of missing important findings in hypothesis testing.

[Hypothesis Testing Guide for Data Science Beginners (analyticsvidhya.com)](https://www.analyticsvidhya.com/blog/2021/07/hypothesis-testing-made-easy-for-the-data-science-beginners/)

[Hypothesis Testing - A Deep Dive into Hypothesis Testing, The Backbone of Statistical Inference - Machine Learning Plus](https://www.machinelearningplus.com/statistics/hypothesis-testing/)

[Understanding Hypothesis Testing in Data Science: T-tests, F-tests, and More | by Mark Stent | Medium](https://medium.com/@markstent/understanding-hypothesis-testing-in-data-science-t-tests-f-tests-and-more-f520cce06f69)