# Hypothesis Testing

Hypothesis testing is a statistical framework uses to make inferences about populations based on sample data. Its more theoretical in a way. When it comes to real life application part mostly A/B testing is used, both are different in a way not completely, one way to put is as Hypothesis testing involves broader range of statistical tests to make decisions or make inferences including, but not limited to A/B testing.

My understanding is that typically in EDA process, after checking the correlation between variables, hypothesis testing can be done to check its statistical significance, to provides statistical evidence for the claim.

Applications: Marketing, Social Sciences, Sports Analytics, Pharmaceuticals etc…

Terminologies:

**Confusion Matrix:** The confusion matrix tells us how accurate ML algorithm predicts. The size of the confusion matrix is predicted by the no of variables we predict.

**False Positive:** If the predicted value is True, but the actual value is False. Getting a small P-value when there is no difference is called false positive. Smaller P-value does not imply the affect size. Also known as **Type 1 error.** (we reject the null hypothesis when it is true)

**False Negative:** If the predicted value is False, but the actual value is True. Also known as **Type 2 error.** (we fail to reject the null hypothesis when it is false)

This process involves few steps:

1. **Defining the problem statement** or statement that we want to check is either correct or not.

Ex: Does studying hours impact the student’s marks?

1. **Then defining the Null and Alternate Hypothesis**.

Null Hypothesis(H0): Studying hours has **NO** impact on student’s marks

Alternative Hypothesis(H1): Studying hours has an impact on student’s marks.

**NOTE:** Null Hypothesis(H0): We pick the null hypothesis which is easier to simulate. (or)

The hypothesis that there no difference between the things is called as Null Hypothesis.

1. **Set the level of significance level**

Generally it is decided by domain experts, but mostly used is 5% which is 0.05. It depends on the application like if we checking for medical experiment, we can’t afford to have more **False Positives** so threshold should be small**,** but if we want to check if an ice cream truck arrives on time or not then we can consider 0.2 as threshold also. That suggests we can afford to have False positives 2 out of ten times.