HYPOTHESIS TESTING AND P-VALUE

Hypothesis
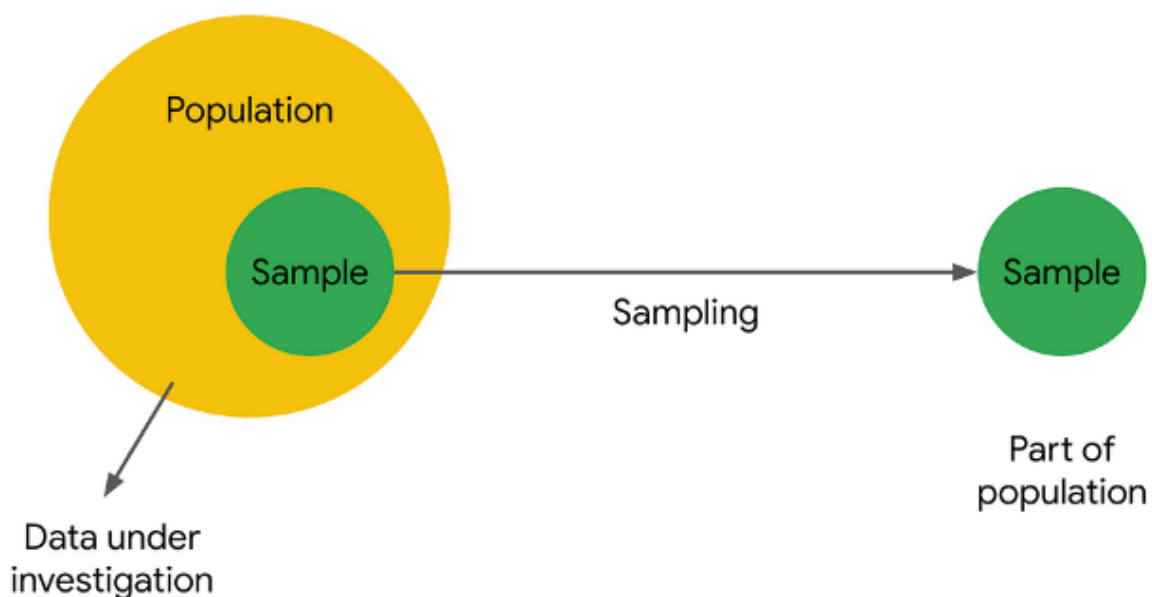

At first Understand,

Population -→ (It is a Full Collection). The population refers to the entire group or set of individuals, objects, or measurements being studied.

Sample→The sample is a subset of the population that is selected to make inferences about the population.
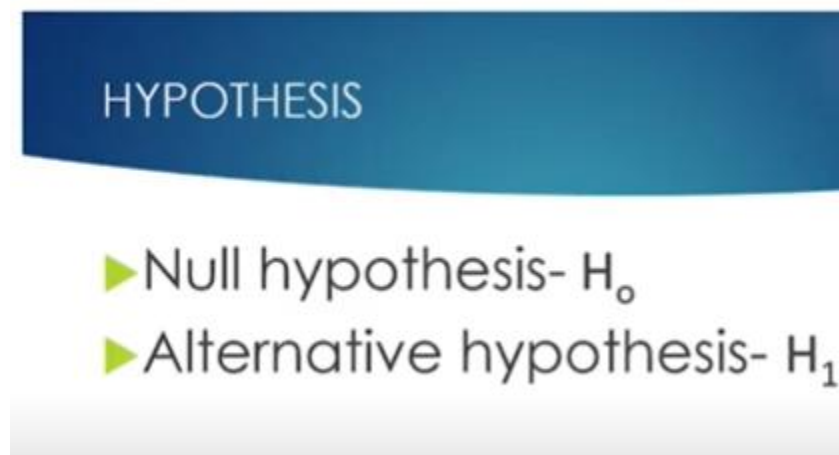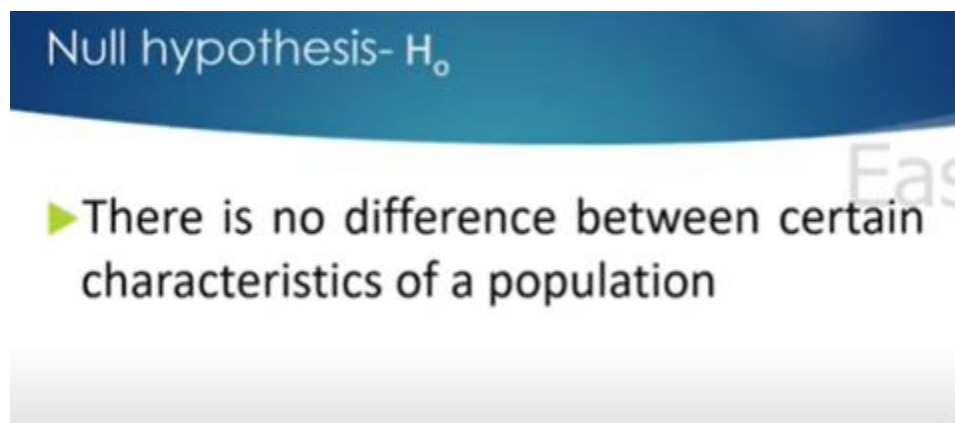
Population---→ Parameters

Sample-→ Statistic

For Example, I cannot take all the **population** to research for some calculation. So I Taking Some **Sample** from that population and Analyzing for that Particular Calculation.

Testing Of Hypothesis

First I Give Assumption, then Testing, at last Final Decision of Hypothesis.

Null Hypothesis

Alternate Hypothesis

Alternative hypothesis- H_1

▶There is difference between certain characteristics of a population

How to Know (Understand) the P-value

P- value (significant at 5% or 0.05)

▶ A p-value is used in hypothesis testing to help you **accept** or **reject** the null hypothesis. The p-value is the **evidence against** a null hypothesis. The **smaller** the p-value, the **stronger** the evidence that you should **reject** the null hypothesis.
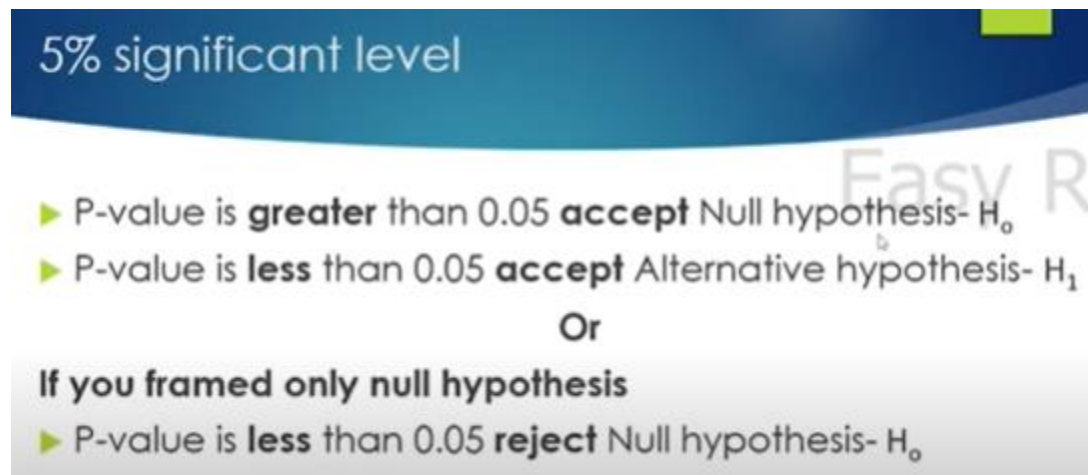
Categories

Significant levels and confidence levels

- Significant level 1% or confidence level 99%(P-value=0.01)
- Significant level 5% or confidence level 95% (P-value=0.05)
- Significant level 10% or confidence level 90%(P-value=0.10)

Error level→1%, Confidence Level→99% same

Error level→5%, Confidence Level→95%

Error level→10%, Confidence Level→90%

If 5%Significant Level

5% significant level

- P-value is **greater** than 0.05 **accept** Null hypothesis- H_o
- P-value is **less** than 0.05 **accept** Alternative hypothesis- H_1

Or

If you framed only null hypothesis

- P-value is **less** than 0.05 **reject** Null hypothesis- H_o

5% Significance Level In Visualization(Acceptance and Rejection Regions)

The Left Hand side and right hand side (0.25+0.25)=0.5 . It is a Error Level and called as Rejection region.

The Inner side of (0.475+0.475)=0.95 called Acceptance Region

Acceptance and rejection regions in case of a two-tailed test (with 5% significance level)
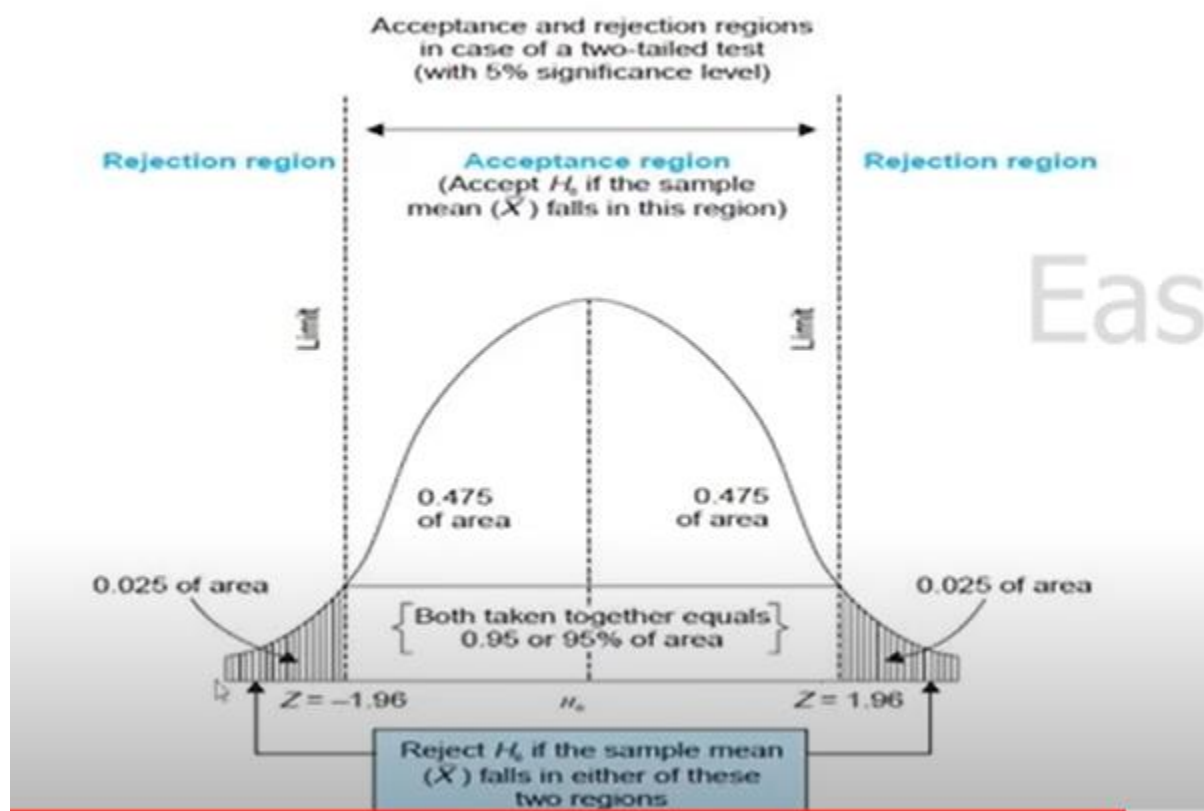
Summary

hypothesis testing in Python is a powerful tool that enables you to draw reliable, statistically valid conclusions from your data, test assumptions, compare groups,

and make informed decisions - all of which are essential for data analysis, machine learning, and scientific research.

UnPaired T-Test

```python
from scipy.stats import ttest_ind
dataset=dataset.dropna()
male = dataset[dataset['gender']=='M']['salary']
female = dataset[dataset['gender']=='F']['salary']
ttest_ind(male, female)
```

TtestResult(statistic=2.0937842507001165, pvalue=0.03746230397067219, df=213.0)

The t-statistic of 2.0937842507001165 suggests that there is a difference between the two groups, and the p-value of 0.03746230397067219 suggests that there is a relatively low probability of observing a t-statistic as extreme as the one calculated if the null hypothesis is true. If the significance level is set to 0.05, then the null hypothesis would be rejected, and it would be concluded that there is a statistically significant difference between the two groups

p-value=0.03 (0.03<0.05) .So the Null Hypothesis is rejected and Accept Alternate Hypothesis

Paired T test

Dependant Sample-Paired T_Test

Same Group(Male) but Different Condition(ssc_p,hsc_p)

```
from scipy.stats import ttest_rel
#dataset=dataset.dropna()
male = dataset[dataset['gender']=='M']['ssc_p']
male1 = dataset[dataset['gender']=='M']['hsc_p']
ttest_rel(male, male1)
```

TtestResult(statistic=0.6842940392688153, pvalue=0.4949370716000775, df=138)

In the given search results, the T-test result is TtestResult(statistic=0.6842940392688153, pvalue=0.4949370716000775, df=138). This means that the statistic value is 0.684, the p-value is 0.4949, and the degrees of freedom is 138. Based on this result, the null hypothesis of no significant difference between the means of the two groups cannot be rejected, as the p-value is greater than the significance level of 0.05.

p-value=0.49 (0.49>0.05) .So accept the Null Hypothesis.