**Hypothesis Testing**

Statistical Hypotheses testing is a formal means of choosing between two distributions on the basis of a particular statistic or random variable generated from one of them

**Neyman-Pearson Paradigm:** There are two types of hypotheses, simple ones where the hypothesis completely species the distribution.

* Null hypothesis H0
* Alternate hypothesis Ha or H1

**The Null Hypothesis:**

An assumption which is approval to the statement is called a null hypothesis. It is an assertion that we hold as true unless we have sufficient statistical evidence to conclude otherwise.

For example- a null hypothesis might assert that the population mean is equal to 100.Unless we obtain sufficient evidence that it is not 100, we will accept it as 100.

So, we write the null hypothesis compactly as H0: µ=100

**Alternate hypothesis:**

The alternative hypothesis is against of the null hypothesis

For the null hypothesis 100, the alternative hypothesis is µ ≠100. We will write it as H1: µ ≠100

Using the symbol H1 to denote the alternative hypothesis, because the null and alternative hypotheses assert exactly opposite statements, only one of them can be true. Rejecting one is equivalent to accepting the other.

***Type I and Type II Errors***

In the context of statistical hypothesis testing, rejecting a true null hypothesis is

Known as a **type I error** and accepting a false null hypothesis is known as a **type II error.**

|  |  |
| --- | --- |
| No error | Type II error |
| Type I error | No error |

**H0 True H0 False**

**Accept H0**

**Reject H0**

**P-value:**

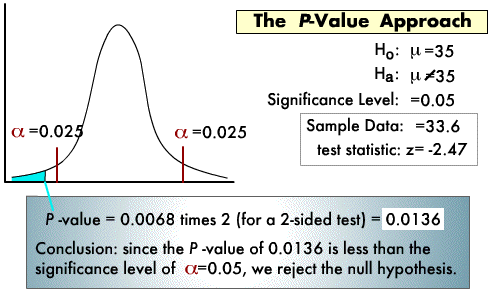
p value is used in hypothesis testing to help support or reject the null hypothesis. The p value is evidence against a null hypothesis. For example,  p value of 0.05 is 5%. This means there is a 5% chance your results could be random. Therefore, the smaller the p-value, the more “significant” to your results.

***The Significance Level***

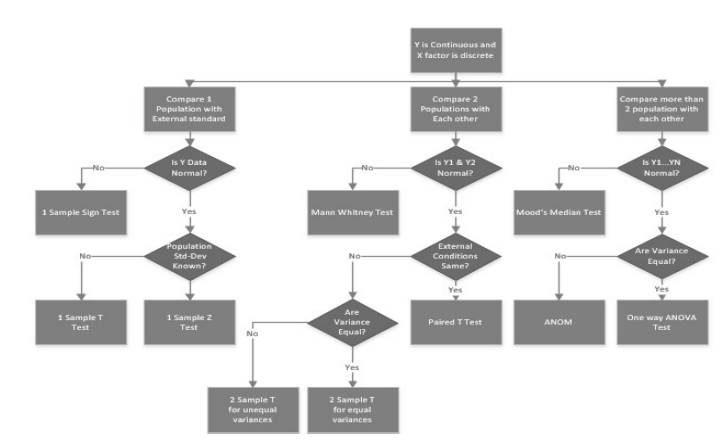
The most common policy in statistical hypothesis testing is to establish a significance level, denoted by α, and to reject H0 when the *p-*value falls below it. When this policy is continued, one can be sure that the maximum probability of type I error is α.

*Rule:* *p-*value is less than α, reject H0.

The standard values for α are 1%, 5%, and 10%. Suppose α is set at 5%. This means that whenever the *p-*value is less than 5%, H0 will be rejected.

****

* Different tests based on type of input and output data types and number of data types



**Mood’s Median Test:**

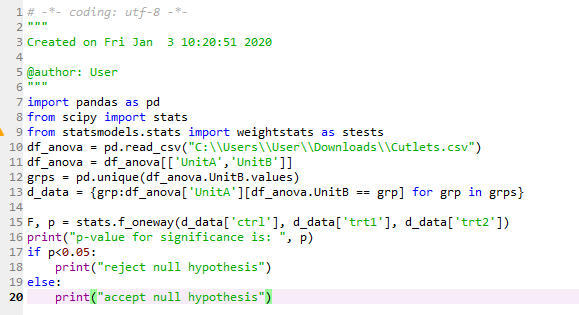
* This test handles outliers well
* It is a nonparametric test that is used to test the equality of medians from two or more populations.
* It provides a nonparametric to alternative to the one-way ANOVA.
* This test works when the Y variable is continuous, discrete-ordinal or discrete-count, and the X variable is discrete with two or more attributes

**One way ANOVA:**

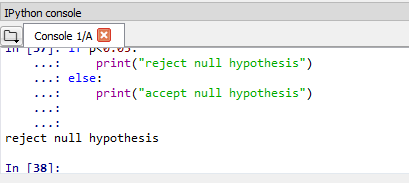
* ANOVA can be used to test equality of means when there are more than 2 populations
* ANOVA can be used with one or two factors.
* Variance should be equal in this test, if variances are not equal then go to ANOM(Analysis of means ).
* If only one factor is varying, then we would use a one-way ANOVA

**For example**:  there are three different categories of plant and their weight and need to check whether all three group are similar or not.

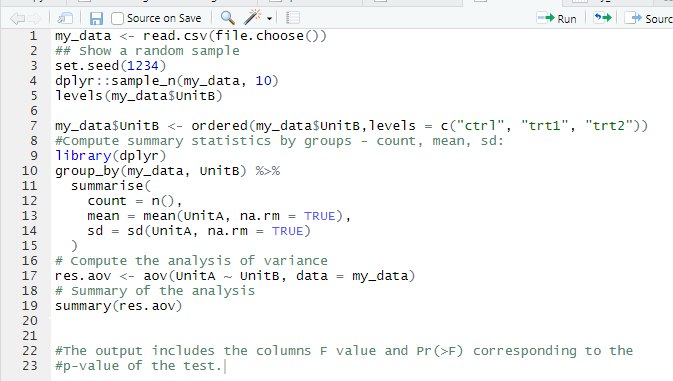
**Python code:**



**Output:**



**R code:**



**Output:**

