Hypothesis Testing

Role of Statistics

- The role of statistics in research is to function as a tool in designing research, analysing its data and drawing conclusions therefrom.
- Most research studies result in a large volume of raw data which must be suitably reduced so that the same can be read easily and can be used for further analysis

Role of Statistics

- The important statistical measures that are used to summarise the survey/research data are:
  (1) measures of central tendency or statistical averages
  (2) measures of dispersion;
  (3) measures of asymmetry (skewness);
  (4) measures of relationship

Role of Statistics

- Amongst the measures of central tendency, the three most important ones are the arithmetic average or mean, median and mode

WHAT IS A HYPOTHESIS?

- Ordinarily, when one talks about hypothesis, one simply means a mere assumption or some supposition to be proved or disproved.
- But for a researcher hypothesis is a formal question that he intends to resolve.

WHAT IS A HYPOTHESIS?

For example,
consider statements like the following ones:
“Students who receive counselling will show a greater increase in creativity than students not receiving counselling”

- Or

- “the automobile A is performing as well as automobile B.” These are hypotheses capable of being objectively verified and tested.

- Thus, we may conclude that a hypothesis states what we are looking for and it is a proposition which can be put to a test to determine its validity

Hypothesis Testing

- **Hypothesis Testing** could be used to interpret and draw conclusions about the population using sample data.

- A **Hypothesis Test** helps in making a decision as to which mutually exclusive statement about the population is best supported by sample data.

Hypothesis Testing

- **Null Hypothesis (H_0) –** It is a statement that is commonly accepted
- It is assumed that the observed result is due to the chance of factor. It is denoted by H_0.

- If it is a test of means then we say that **H_0: $\mu_1 = \mu_2$** , which states that there is no significant difference in the 2 population means.

Hypothesis Testing

- **Alternate Hypothesis(H_1 or H_a) –** As previously mentioned that Null Hypothesis and Alternate Hypothesis are mutually exclusive statements.

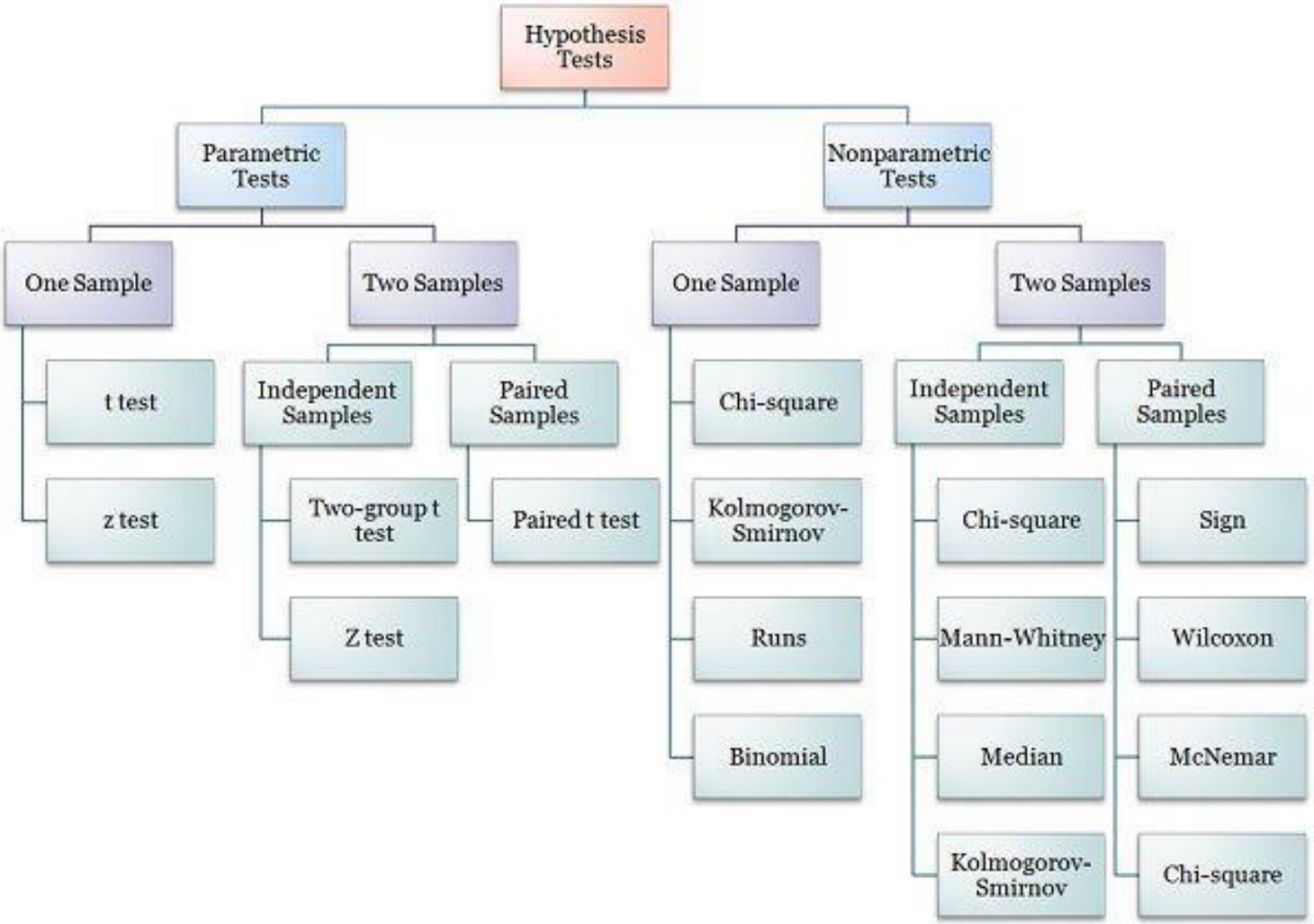
- So if the Null Hypothesis is commonly accepted facts then the Alternate Hypothesis is a real fact-based on observation from the sample data.

- It is denoted by H_1 or H_a. If it is a test of means then we say that $\mathbf{H_1 : \mu_1 \neq \mu_2}$, which states that there is a significant difference in 2 population means.

Hypothesis Testing

	Decision	
	Accept H_0	Reject H_0
H_0 (true)	Correct decision	Type I error (α error)
H_0 (false)	Type II error (β error)	Correct decision

Hypothesis Testing



Hypothesis Testing -Types

- **A parametric test** is used when we make a specific assumption about the underlying distribution of the population from which the sample is being drawn, and which is being investigated.
- **A nonparametric test** is a hypothesis test that does not require the population's distribution to be characterized by certain parameters.
- For example, many hypothesis tests rely on the assumption that the population follows a normal distribution with parameters μ and σ.
- Nonparametric tests do not have this assumption, so they are useful when your data are strongly non-normal

- **Description:**
- **t Test:** Hypothesis testing for small sample size. **Z test:** Hypothesis testing for Large sample. **Chi-square Test:** Test of Significance to determine the difference observed and expected frequencies of certain observations.

- **Critical Region –** The critical region is defined as the region of values in distribution that leads to the rejection of the null hypothesis at some given probability level.
- **One-Tailed Test –** A one-tailed test is a statistical hypothesis test in which the critical area of distribution is either greater than or less than a certain value, but can't be both. For this the alternate hypothesis formulation is $\mathbf{H_1 : \mu_1 > \mu_2}$ **or** $\mathbf{H_1 : \mu_1 < \mu_2}$ **.**
- **Two-Tailed Test –** A two-tailed test is a statistical hypothesis test in which the critical area of distribution is on either of the sides. It tests whether the sample means of 2 or more populations are unequal (in the test of means). For this alternate hypothesis, the formulation is $\mathbf{H_1 : \mu_1 \neq \mu_2}$.

- In either of the above 2 tests if the sample tested falls in the critical region than the alternate hypothesis holds to be true and the null hypothesis is rejected. The alternate hypothesis is made as a conclusive observation for the population-based on sample data.

Error terms in hypothesis testing

- **Type-I error –** This error occurs when insights drawn from sample data lead to rejection of the null hypothesis even when it is true. This error could be controlled as it has direct bearing with a **level of significance.**
- **Type-II error –** This error occurs when insights from sample data result in failing to reject the null hypothesis although it is false.

- **Level of Significance –** It is the probability of making type I error and is denoted by **α.** It is the maximum probability of making type I error. As per standard for 95% confidence level value of alpha is 0.05. This means that there is a 5% probability of making type I error or rejecting the null hypothesis even when it is true.

- **What Are Degrees of Freedom?**
- Degrees of freedom refers to the maximum number of logically independent values, which are values that have the freedom to vary, in the data sample.
- The formu
the size of

The formula for degrees of freedom equals the size of the data sample minus one:

$$D_f = N - 1$$

where:

D_f = degrees of freedom

N = sample size

- Formulate the Null and Alternate Hypothesis
- Based on data and probability distribution select the hypothesis test to be performed
- Based on the business are and problem statement selects the level of significance if 0.05 (standard alpha) is not acceptable.
- Calculate test statistics on the sample data collected
- Calculate the p-value
- Based on p-value draw insights to reject or fail to reject the null hypothesis.
- Draw your business conclusion.

- For Null hypothesis H_0: $\mu \geq \mu_0$ vs alternative hypothesis H_1: $\mu < \mu_0$, it is lower/left-tailed (one tailed).
- For Null hypothesis H_0: $\mu \leq \mu_0$ vs alternative hypothesis H_1: $\mu > \mu_0$, it is upper/right-tailed (one tailed).
- For Null hypothesis H_0: $\mu = \mu_0$ vs alternative hypothesis H_1: $\mu \neq \mu_0$, it is two-tailed.
- Third, calculate the standard score : $Z = \frac{(\bar{X} - \mu_0)}{\sigma}$,

- **What Are Degrees of Freedom?**
- Degrees of freedom refers to the maximum number of logically independent values, which are values that have the freedom to vary, in the data sample.
- The formu
the size of

The formula for degrees of freedom equals the size of the data sample minus one:

$$D_f = N - 1$$

where:

D_f = degrees of freedom

N = sample size

- **Description:**
- **t Test:** Hypothesis testing for small sample size. **Z test:** Hypothesis testing for Large sample. **Chi-square Test:** Test of Significance to determine the difference observed and expected frequencies of certain observations.