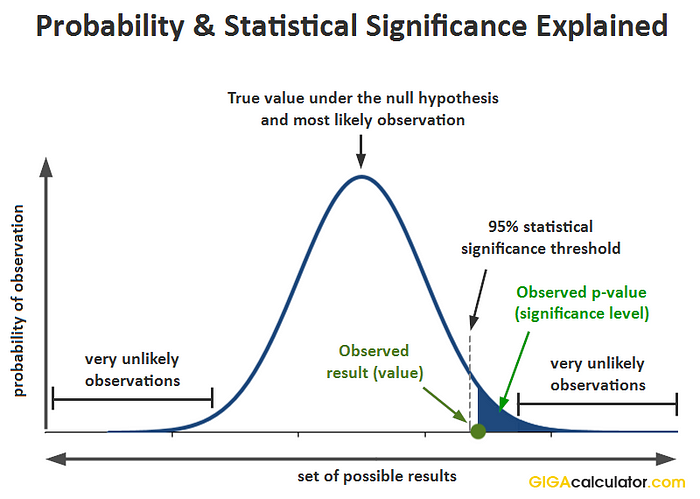
**Using Hypothesis Testing such as Z-test, P-Value, T-test, Chi-Square test, ANOVA.**

**What should be the strategy, when to use what test?**

Before proceeding to all the tests, let’s have a brief understanding of Hypothesis Testing:

Hypothesis testing is a tool for making statistical inferences about the population data. It is an analysis tool that tests assumptions and determines how likely something is within a given standard of accuracy. Hypothesis testing provides a way to verify whether the results of an experiment are valid.

source: GIGAcalculator.com

**Null Hypothesis (Ho):** The null hypothesis states that the two samples of the population are the same.

**Alternate Hypothesis (Ha):** It states that there is a significant difference between the two samples of the population.

For the null hypothesis, the same means are assumed to be equal, and we have H0: µ1= µ2

& for the alternate hypothesis, the sample means are unequal, and we have Ha: µ1≠ µ2.

**Questions on Null Hypothesis:**

Question 1: A medical experiment and trial is conducted to check if a particular drug can serve as the vaccine for Covid-19 and can prevent from occurrence of Corona. Write the null hypothesis and the alternate hypothesis for this situation.

Solution:

The given situation refers to a possible new drug and its effectiveness of being a vaccine for Covid-19 or not. The null hypothesis (Ho) and alternate hypothesis (Ha) for this medical experiment is as follows.

Ho: The use of the new drug serves as a vaccine and helps for the prevention of Covid-19.

Ha: The use of the new drug is not helpful for the prevention of Covid-19.

Example 2: The teacher has prepared a set of important questions and informs the student that preparing these questions helps in scoring more than 60% marks in the board exams. Write the null hypothesis and the alternate hypothesis for this situation.

Solution:

The given situation refers to the teacher who has claimed that her important questions help to score more than 60% marks in the board exams. The null hypothesis (Ho) and alternate hypothesis (Ha) for this situation is as follows.

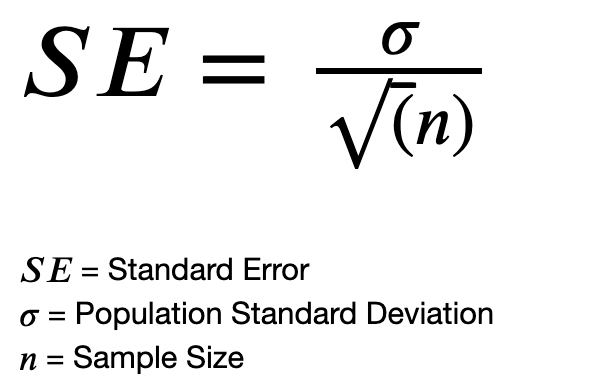
Ho: The important questions given by the teacher are helpful for the students to score more than 60% marks in the board exams.

Ha: The important questions given by the teacher do not really help the students to get a score of more than 60% in the board exams.

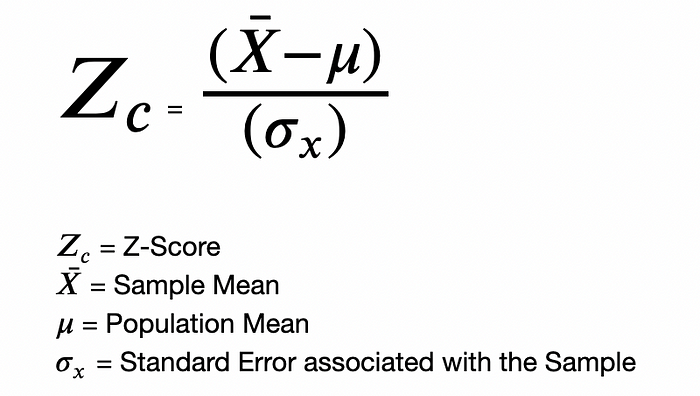
**Hypothesis Testing P Value:**

* In hypothesis testing, the [p value](https://www.cuemath.com/p-value-formula/) is used to indicate whether the results obtained after conducting a test are statistically significant or not.
* Basically, it decides whether we should accept our Null Hypothesis or reject it. The lower the p-value, the more surprising the evidence is, the more ridiculous our null hypothesis looks. And when we feel ridiculous about our null hypothesis, we simply reject it and accept our Alternate Hypothesis.
* If we found the p-value is lower than the predetermined significance value (often called alpha or threshold value (The alpha level can be defined as the acceptable risk of incorrectly rejecting the null hypothesis. The alpha level is usually chosen between 1% to 5%.)) then we reject the null hypothesis. The alpha should always be set before an experiment to avoid bias.
* For example, we generally consider a large population data to be in Normal Distribution so while selecting alpha for that distribution we select it as 0.05 (it means we are accepting if it lies in the 95 percent of our distribution). This means that if our p-value is less than 0.05 we will reject the null hypothesis.

**Steps1: C**alculate the [Standard Error](https://en.wikipedia.org/wiki/Standard_error) of the sample, which is the Population Standard Deviation divided by the square root of sample size (n).



**Step2**: After finding the Standard Error, we take a sample and the mean of that sample and then find the Z-score associated with that mean value.

 Z = (Sample Mean - Population Mean)/ (Standard Error)

**Step3:** After the p-value associated with the Z-score is calculated, we refer the Z-table to find the probability of the Z-score calculated. Then, to find the p-value, we subtract that probability from 1.

P-Value = 1 - Probability(Z-score)

**Step4:** Finally, we check if the calculated p-value is greater than the significance level or not.

If the ***P-value > Significance Level***, then we ***Fail to Reject the Null Hypothesis****.*  
Or else, if the ***P-value < Significance Level***, we ***Reject the Null Hypothesis****.*

**Questions on P-Value:**

Example 1:A statistician is testing the hypothesis H0: μ = 120 using the approach of alternative hypothesis Hα: μ > 120 and assuming that α = 0.05. The sample values that he took are as n =40, σ = 32.17 and x̄ = 105.37. What is the conclusion for this hypothesis?

A picture containing diagram

Description automatically generatedSolution:

We know that:

On substitution: 32.17/40 = 5.0865

As per the test static formula, we get.

t = (105.37 – 120) / 5.0865

t = -2.8762

Using the Z-Score table, finding the value of P(t > -2.8762), we get,

P (t < -2.8762) = P (t > 2.8762) = 0.003

If,

P (t > -2.8762) =1 - 0.003 =0.997

P- value =0.997 > 0.05

As the value of p > 0.05, the null hypothesis is accepted.

Therefore, the **null hypothesis is accepted**.

**Example 2: P-value is 0.3105. If the level of significance is 5%, find if we can reject the null hypothesis.**

**Solution:** Looking at the P-value table, the p-value of 0.3105 is greater than the level of significance of 0.05 (5%), we fail to reject the null hypothesis.

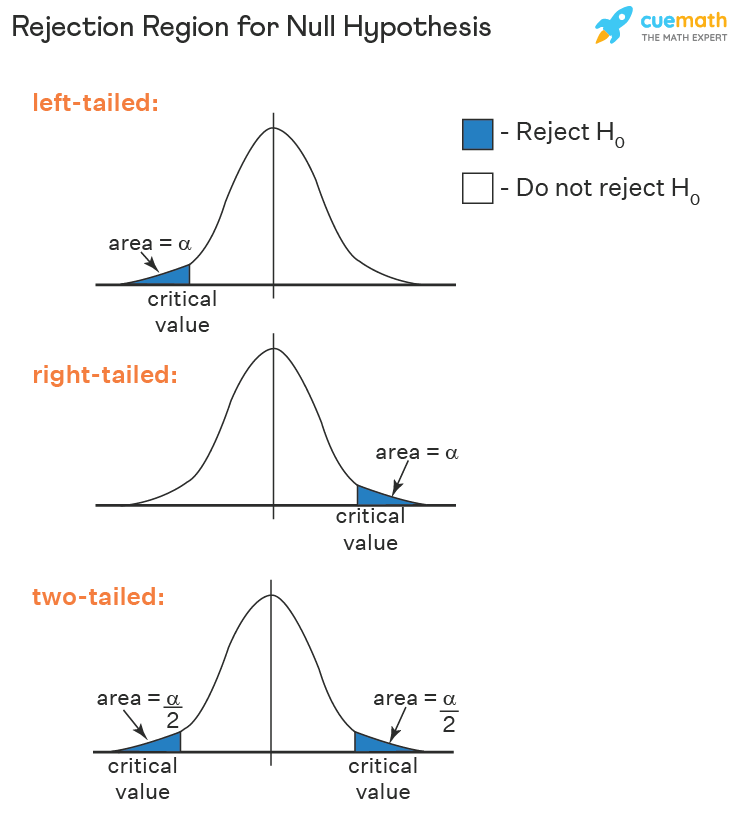
**Hypothesis Testing Z-test:**

* A z-test is a test that is used to check if the [means](https://www.cuemath.com/data/mean/) of two populations are different or not provided the data follows a normal distribution.
* To start with, the null hypothesis and the alternative hypothesis must be set up and the value of the z test statistic must be calculated. The decision criterion is based on the z critical value.

**Conditions to apply z-test:**

* Z test is a statistical test that is conducted on normally distributed data to check if there is a difference in means of two data sets.
* The sample size should be greater than 30 and the population variance must be known to perform a z test.
* The one-sample z test checks if there is a difference in the sample and population mean,
* The two sample z test checks if the means of two different groups are equal.

Refer: https://www.z-table.com/



**Question 1:** A teacher claims that the mean score of students in his class is greater than 82 with a standard deviation of 20. If a sample of 81 students was selected with a mean score of 90 then check if there is enough evidence to support this claim at a 0.05 significance level.

**Solution:**

As the sample size is 81 and population standard deviation is known, this is an example of a right-tailed one-sample z test.

Ho: μ=82

Ha: μ>82

From the z table the critical value at α = 1.645

Graphical user interface

Description automatically generated

x¯ = 90, μ = 82, n = 81, σ = 20

z = 3.6

As 3.6 > 1.645 thus, the null hypothesis is rejected, and it is concluded that there is enough evidence to support the teacher's claim.

**Answer:** Reject the null hypothesis

**Question 2:** An online medicine shop claims that the mean delivery time for medicines is less than 120 minutes with a standard deviation of 30 minutes. Is there enough evidence to support this claim at a 0.05 significance level if 49 orders were examined with a mean of 100 minutes?

**Solution:** As the sample size is 49 and population standard deviation is known, this is an example of a left-tailed one-sample z test.

Ho: μ=120

Ha: μ<120

From the z table the critical value at α = -1.645. A negative sign is used as this is a left tailed test.

Graphical user interface

Description automatically generated

x¯ = 100, μ = 120, n = 49, σ = 30

z = -4.66

As -4.66 < -1.645 thus, the null hypothesis is rejected, and it is concluded that there is enough evidence to support the medicine shop's claim.

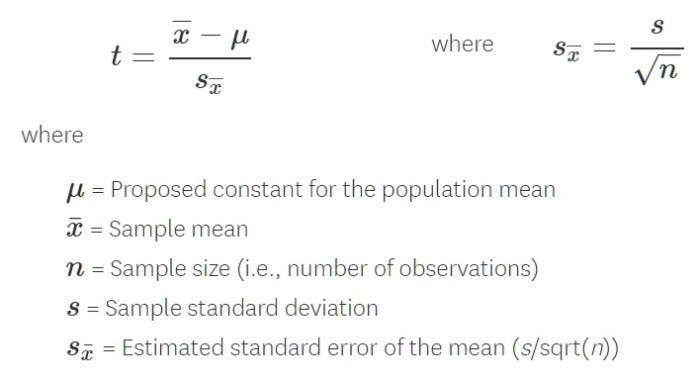
**Answer:** Reject the null hypothesis

**Hypothesis Testing T-test:**

* The t-test formula helps us to compare the average values of two data sets and determine if they belong to the same population or are they different.
* The t-score is compared with the critical value obtained from the t-table. The large t-score indicates that the groups are different, and a small t-score indicates that the groups are similar.

There are three different versions of t-tests:

* One sample t-test which tells **whether means of sample and population are different**.



* Two sample t-test also is known as independent t-test — it compares the means of two independent groups and determines whether there is statistical evidence that the associated population means are significantly different.

Diagram

Description automatically generated

* Paired t-test when you want to compare means of the different samples from the same group or **which compares means from the same group at different times***.*

**Condition to apply t-test:**

* The sample size is smaller than 30 and the population variance is not known to perform a t-test.

**Question 1:** A company wants to improve its sales. The previous sales data indicated that the average sale of 25 salesmen was $50 per transaction. After training, the recent data showed an average sale of $80 per transaction. If the standard deviation is $15, find the t-score. Has the training provided improved the sales?

**Solution:**

Ho: The population mean = the claimed value⇒μ = μ0

Ha: The population mean not equal to the claimed value⇒μ ≠ μ0

t**-**test formula for independent test is t= x (sample mean) − μ/ s√n

Mean sale = 80, μ = 50, s= 15 and n= 25

substituting the values, we get t= (80-50)/(15/√25)

t = (30 ×5)/10 = 10

Looking at the t-table we find 10 > 1.711. (I.e., CV for α = 0.05). ∴ the accepted hypothesis is not true. Thus, we conclude that the training boosted the sales.

**Hypothesis Testing Chi-Square Test:**

* The Chi-Square test is used when we perform**hypothesis testing on two categorical variables**from a single population or we can say that to**compare categorical variables**from a single population.
* Text, letter

  Description automatically generatedBy this we find is there any significant association between the two categorical variables.

The hypothesis being tested for chi-square is:

**Null**: Variable A and Variable B are independent

**Alternate**: Variable A and Variable B are not independent.

The Chi-Square test gives a P-value to help you know the correlation if any!

A hypothesis is in consideration, that a given condition or statement might be true, which we can test later. For example

* A very small Chi-Square test statistic indicates that the collected data matches the expected data extremely well.
* A very large Chi-Square test statistic indicates that the data does not match very well. If the chi-square value is large, the null hypothesis is rejected.

**Applications:**

* used by Biologists to determine if there is a significant association between the two variables, such as the association between two species in a community.
* used by Genetic analysts to interpret the numbers in various phenotypic classes.
* used in various statistical procedures to help to decide if to hold onto or reject the hypothesis.
* used in medical literature to compare the incidence of the same characteristics in two or more groups.

**Question 1:** Calculate the Chi-square value for the following data of incidences of water-borne diseases in three tropical regions.

Table

Description automatically generated

**Solution:**

Table

Description automatically generatedSetting up the following table:

**Answer:** Chi Square = 125.516

**Hypothesis Testing ANOVA Test:**

* It is also called an analysis of variance and is used to compare multiple (three or more) samples with a single test.
* It is used when the categorical feature has more than two categories.

The hypothesis being tested in ANOVA is:

**Null**: All pairs of samples are same i.e., all sample means are equal

**Alternate**: At least one pair of samples is significantly different

