Hypothesis test - Basics

Reshmi Dey
13/ 11 / 2021

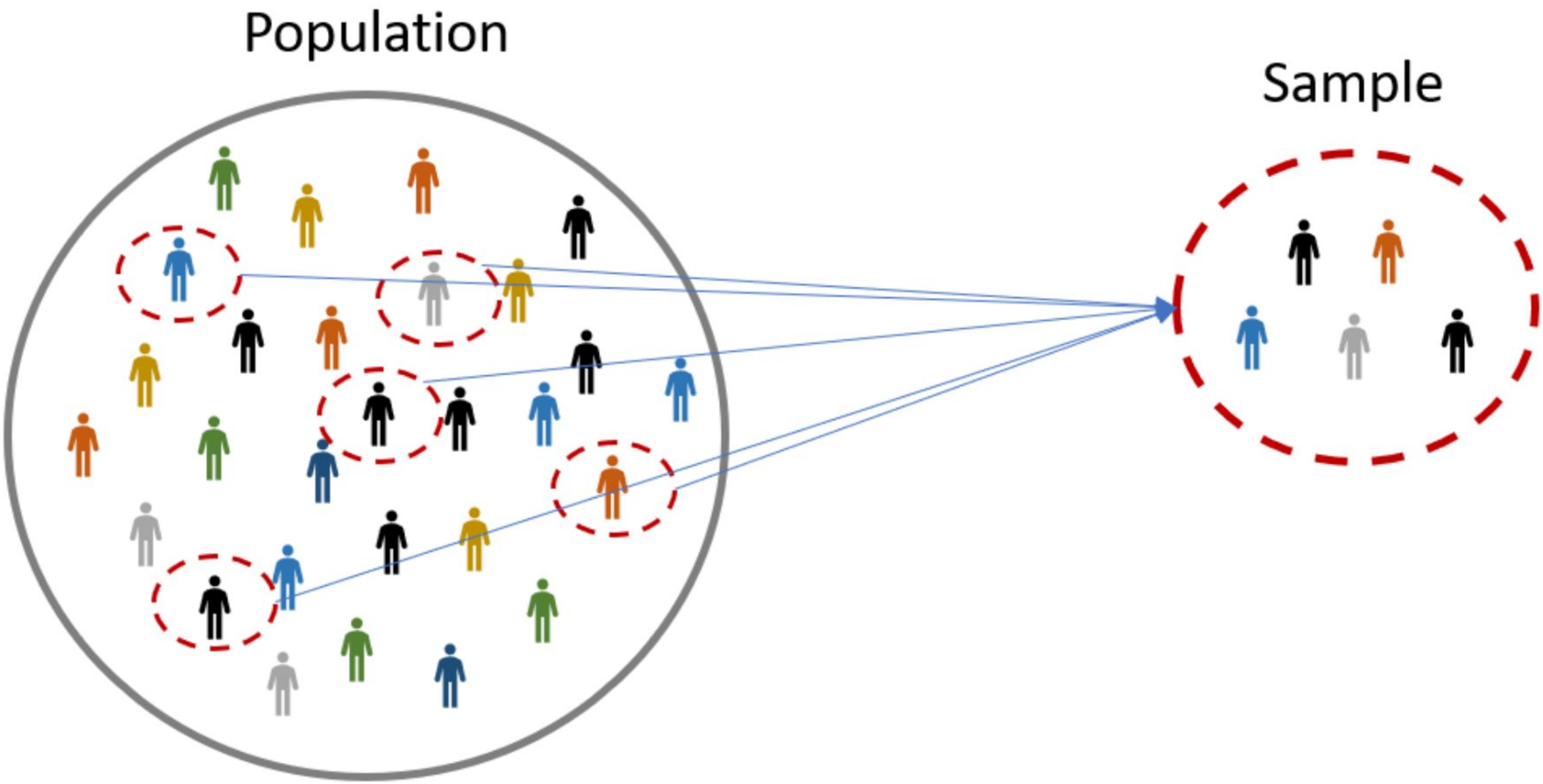
Population And Sample.

What is Population?

In statistics, population is the entire set of items from which you draw data for a statistical study. It can be a group of individuals, a set of items, etc. It makes up the data pool for a study.

What is a Sample?

A sample represents the group of interest from the population, which you will use to represent the data. The sample is an unbiased subset of the population that best represents the whole data.


Population
Sample

Types of Sampling.

(i) Purposive Sampling,

(ii) Random Sampling,

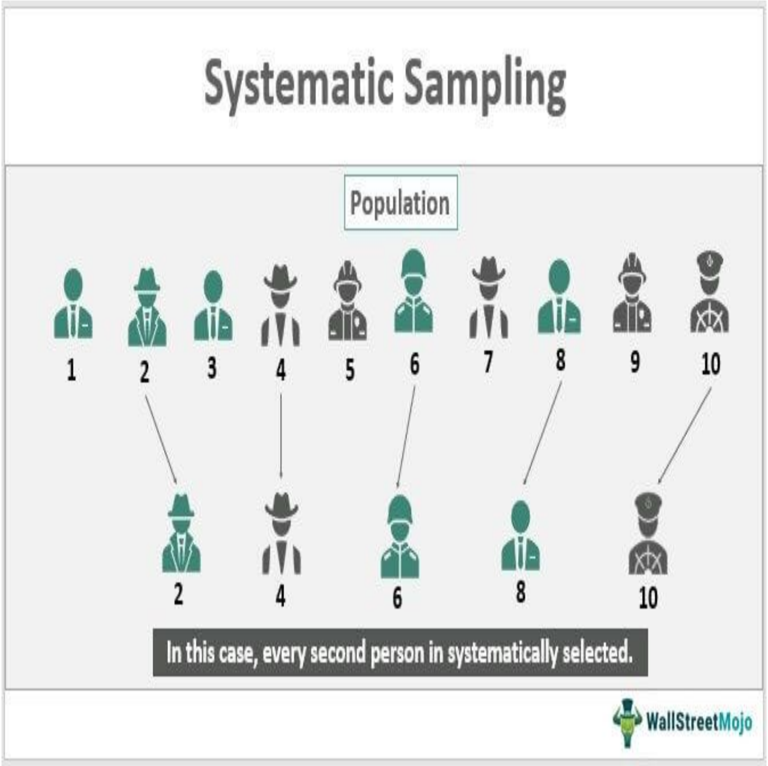
(iii) Stratified Sampling, and

(iv) Systematic Sampling.

(i) Purposive Sampling :- Purposive sampling is one in which the sample units are selected with definite purpose in view. For example if we want to give the picture that the standard of living has increased in the city of New Delhi, we may take individuals in the sample from rich and posh localities and ignore the localities where low income group and the middle class families live. This sampling suffers from the drawbacks of favouritism and nepotism and does give a representative sample of the population.

(ii) Random Sampling :- In this case the sample units are selected at random. A random sample is one in which each unit of population has an equal chance of being included in it. Suppose we take a sample of size 'n' from a finite population of size 'N'. Then there are ${}^{N}C_{n}$ possible samples. A sampling technique in which each ${}^{N}C_{n}$ samples has an equal chance of being selected is known ***random sampling*** and the sample obtained by this technique is termed as ***random sample***.

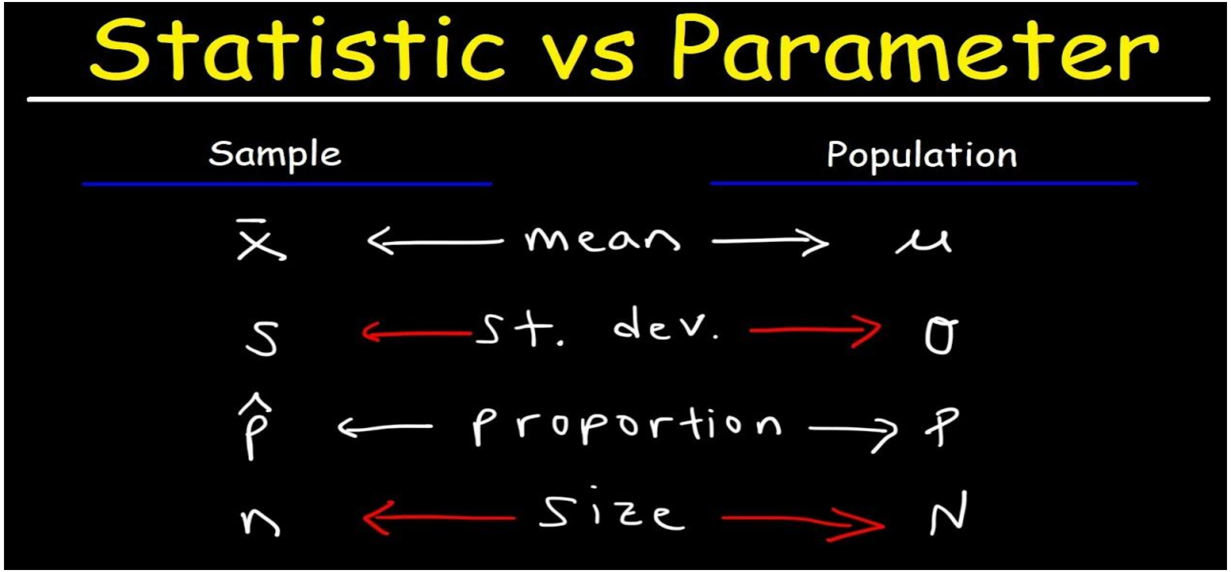
(iii) Stratified Sampling :- Here the population is divided into a number of homogeneous groups, called ***strata***, which differ from one another but each of these groups is homogeneous within itself. The strata should define a partition of the population. That is, it should be collectively exhaustive and mutually exclusive: every element in the population must be assigned to one and only one stratum. Then simple random sampling is applied within each stratum.



(iv) Systematic Sampling :- Systematic sampling is a type of probability sampling method in which sample members from a larger population are selected according to a random starting point but with a fixed, periodic interval. This interval, called the sampling interval, is calculated by dividing the population size by the desired sample size. Despite the sample population being selected in advance, systematic sampling is still thought of as being random if the periodic interval is determined beforehand and the starting point is random.

Parameter and Statistics.

In order to avoid verbal confusion with the statistical constants of the population, which is mean (μ), variance(σ^2),etc., which are usually referred to as ***parameters***, statistical measures computed from the sample observations alone, e.g., mean ($\overline{x}$), variance (s^2), etc., have been termed by Professor R.A. Fisher as ***statistics***.



Standard Error.

The standard deviation of the sampling of a statistic is known as its Standard Error, abbreviated as S.E. The standard errors of some of the well known statistics , for large samples, are given in the following table, where ‘n’ is the sample size, σ^2 the population variance, and P the population proportion, and Q=1-P; n1 and n2 represent the sizes of two independent random samples respectively drawn from the given population (s).

S. No.	Statistic	Standard Error
1.	Sample mean : $\bar{x}$	$\sigma/\sqrt{n}$
2.	Observed sample proportion 'p'	$\sqrt{PQ/n}$
3.	Sample *s.d.* : s	$\sqrt{\sigma^2/2n}$
4.	Sample variance : s^2	$\sigma^2\sqrt{2/n}$
5.	Sample quartiles	$1{\cdot}36263\ \sigma/\sqrt{n}$
6.	Sample median	$1{\cdot}25331\ \sigma/\sqrt{n}$
7.	Sample correlation coefficient (r)	$(1-\rho^2)/\sqrt{n}$, ρ being the population correlation coefficient
8.	Sample moment : μ_3	$\sigma^3\sqrt{96/n}$
9.	Sample moment : μ_4	$\sigma^4\sqrt{96/n}$
10.	Sample coefficient of variation (v)	$\frac{v}{\sqrt{2n}}\sqrt{1+\frac{2v^2}{10^4}} \simeq \frac{v}{\sqrt{2n}}$
11.	Difference of two sample means : $(\bar{x}_1-\bar{x}_2)$	$\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}$
12.	Difference of two sample *s.d.*'s : (s_1-s_2)	$\sqrt{\frac{\sigma_1^2}{2n_1}+\frac{\sigma_2^2}{2n_2}}$
13.	Difference of two sample proportions : (p_1-p_2)	[illegible]

Hypothesis Tests

What is a Hypothesis?

A hypothesis is an **educated guess** about something in the world around you. It should be testable, either by experiment or observation. For example:

- A new medicine you think might work.
- A way of teaching you think might be better.
- A possible location of new species.

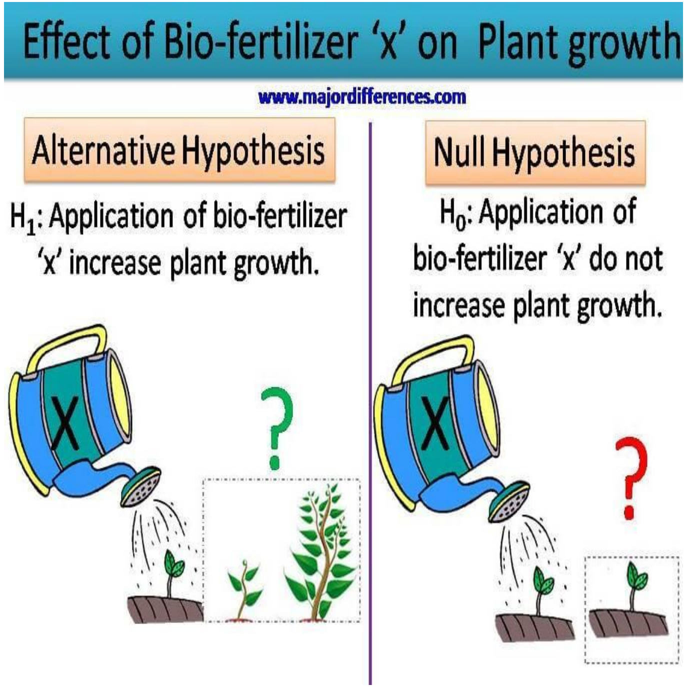
Hypothesis testing in statistics is a way for you to test the results of a survey or experiment to see if you have meaningful results. You're basically testing whether your results are valid by figuring out the odds that your results have happened by chance. If your results may have happened by chance, the experiment won't be repeatable and so has little use.

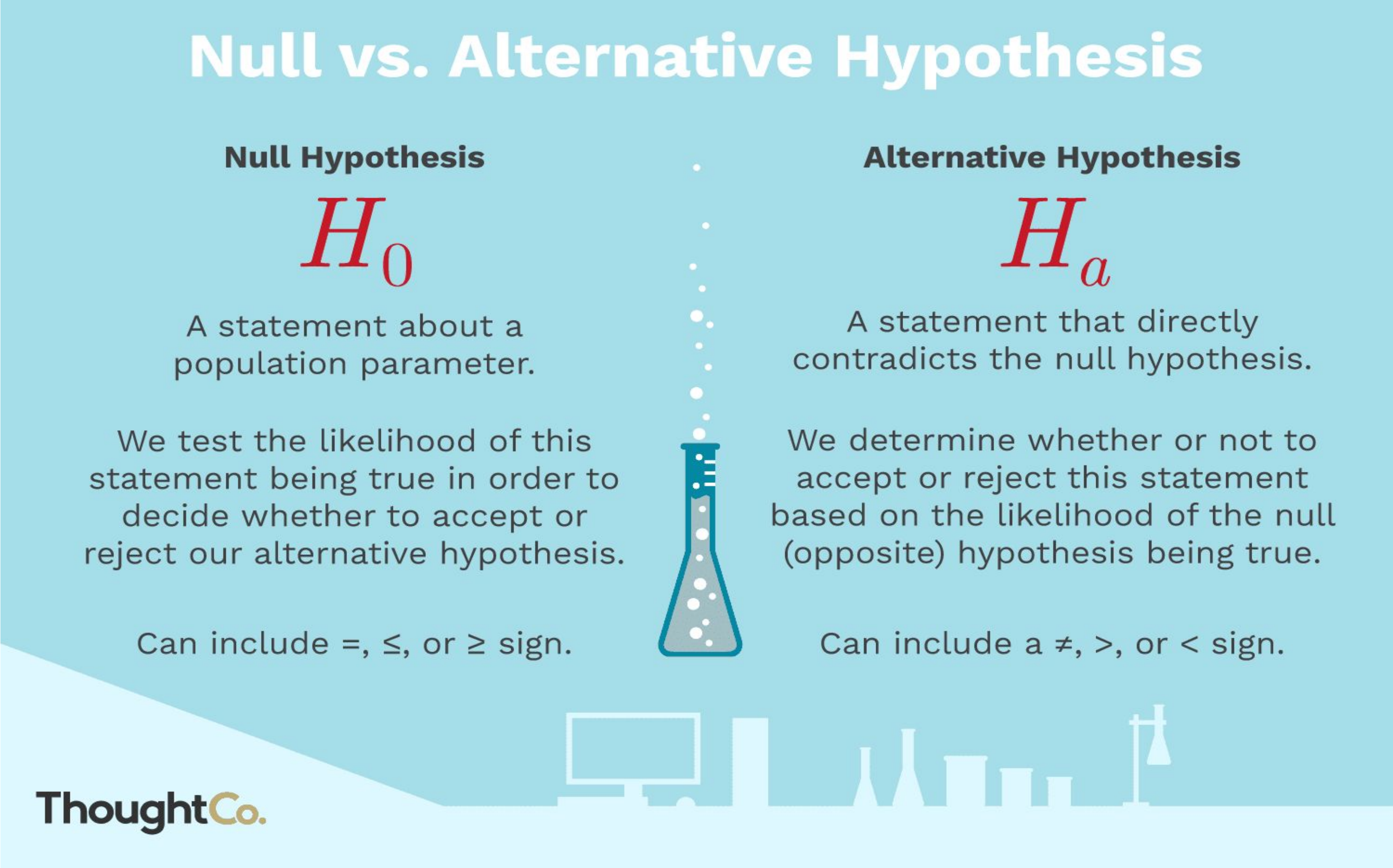
Null and Alternative Hypothesis.

For applying a test we first set up a hypothesis --- a definite statement about the population parameter. Such a hypothesis, which is usually a hypothesis of no difference, is called *Null Hypothesis* and is usually denoted by H_o . Null Hypothesis is the hypothesis which is tested for possible rejection under the assumption that it is true.

Any hypothesis which is complementary to the null hypothesis is called an *Alternative hypothesis*, usually denoted by H_1. For example,if we want to test the null hypothesis that the population has a specified mean μ_o, i.e., $H_o : \mu = \mu_o$, then the alternative hypothesis could be (i) $H_1 : \mu \neq \mu_o$ (ii) $H_1 : \mu > \mu_o$ (iii) $H_1 : \mu < \mu_o$

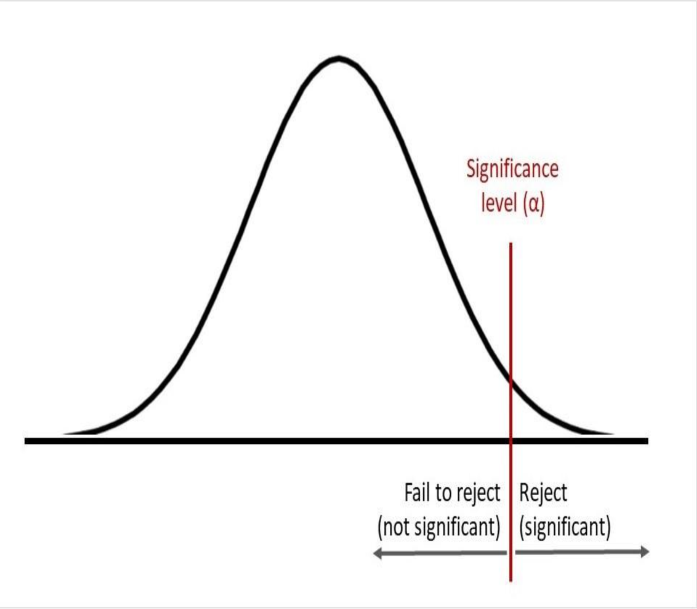
The alternative hypothesis in (i) is known as two tailed alternative and the alternative hypothesis in (ii) and (iii) are known as one tailed alternative.




Null vs. Alternative Hypothesis
Null Hypothesis
H_0
A statement about a population parameter.
We test the likelihood of this statement being true in order to decide whether to accept or reject our alternative hypothesis.
Can include =, ≤, or ≥ sign.
Alternative Hypothesis
H_a
A statement that directly contradicts the null hypothesis.
We determine whether or not to accept or reject this statement based on the likelihood of the null (opposite) hypothesis being true.
Can include a ≠, >, or < sign.
ThoughtCo.

Level of significance.

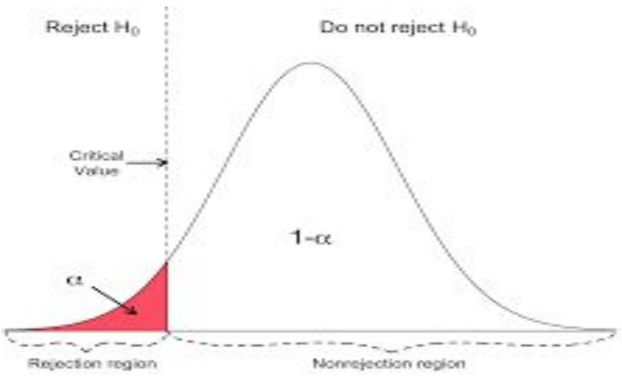
Probability of rejecting a hypothesis when it is actually true is called the level of significance. Basically level of significance is the size of type 1 error. The levels of significance usually employed in testing of hypothesis are 5% and 1%. The level of significance is always fixed in advance before collecting the sample information.

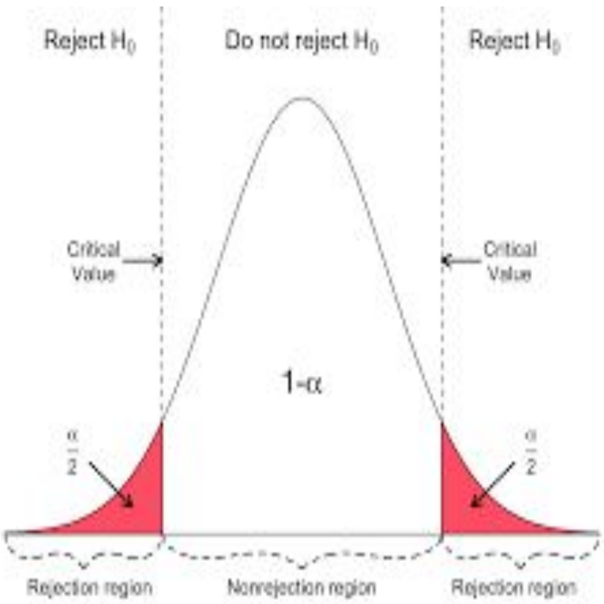


Carsten Grube. Statistical Data Analysis. dataZ4s.com

Critical region.

A region in the sample space S which amounts to the rejection of H_0 is termed as **Critical region** or **Rejection region**. And the complementary set of rejection region is the **Acceptance region.**





Types of Errors.

Whenever we perform any statistical test on the data, there are chances of making error while deciding to accept or reject the lot after examining the sample.

We are liable to commit the following two types of errors :

<u>Type I error</u> : Rejecting the null hypothesis when it is true.

<u>Type II error</u> : Fail to reject the null hypothesis when it is false or accepting H_0 when H_1 is correct

Now, p[Rejecting H_0 when it is true] = $\boldsymbol{\alpha}$

And p[Accepting H_1 when it is false] = $\boldsymbol{\beta}$

Then, α and β are called the **sizes of type I error and type II error**, respectively.

In practice, type I error amounts to rejecting a lot when it is good and type II amounts to accepting the lot when it is bad.

α and β are also called **Producer's risk** and **Consumer's risk** respectively.

When H_0 is correct and we accept it based on statistical test or when we reject H_0 when it is false based on statistical test, it is correct and hence no problem

Type I and Type II Error

Null hypothesis is...	True	False
Rejected	**Type I error** False positive Probability = α	**Correct decision** True positive Probability = 1−β
Not rejected	**Correct decision** True negative Probability = 1−α	**Type II error** False negative Probability = β

Scribbr

Test of significance.

It is a device by means of which we may either accept or reject null hypothesis.

A **test of significance** is a formal procedure for comparing observed data with a claim (also called a hypothesis), the truth of which is being assessed.

• The claim is a statement about a parameter, like the population proportion p or the population mean μ.

• The results of a significance test are expressed in terms of a probability that measures how well the data and the claim agree.

Tests of Significance:

The Four-Step Process :-

1. State the null and alternative hypotheses.

2. Calculate the test statistic.

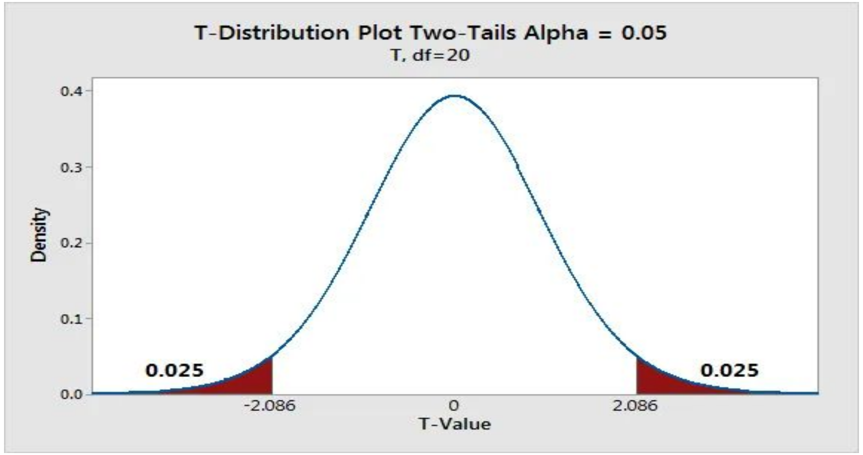
3. Find the P-value.

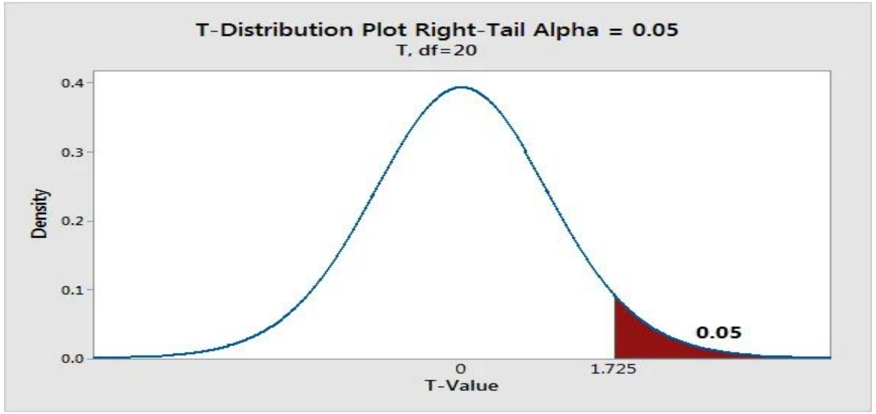
4. Compare P-value with α and decide whether the null hypothesis should be rejected or accepted.

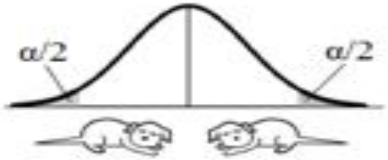
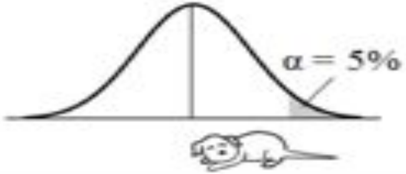
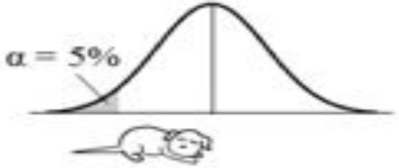
One-tailed and Two-tailed Test

A **one-tailed test** is a statistical test in which the critical area of a distribution is one-sided so that it is either greater than(right tailed) or less than(left tailed) a certain value, but not both.

A **Two-tailed hypothesis tests** are also known as nondirectional and two-sided tests because you can test for effects in both directions. When you perform a two-tailed test, you split the significance level percentage between both tails of the distribution.





Comparison Operator		Tails of the Test	
H_A	H_0		
$\neq$	$=$	2-tailed	
$>$	$\leq$	1- tailed, right-tailed	
$<$	$\geq$	1-tailed, left-tailed	

Example.

Suppose that are two population brands of bulbs, one manufactured by standard process (with mean life μ_1) and the other manufacturer by some new technique (with mean life μ_2).

If we want to if the bulbs differ significantly, then our null hypothesis is H_0: $\mu_1 = \mu_2$ and alternative hypothesis will be H_1:$\mu_1 \neq \mu_2$, thus giving us a two tailed tests.

However, if we want to test if the bulbs produced by new process have higher average life than those produced by standard process, then we have, H_0: $\mu_1 = \mu_2$ and H_0: $\mu_1 < \mu_2$ thus giving us a left-tailed test.

Similarly, for testing if the product of new process is inferior to that of standard process, then we have H_0: $\mu_1 = \mu_2$ and H_0: $\mu_1 > \mu_2$, thus giving us a right-tailed test.

Thus the decision about applying two-tailed or one-tailed(right or left) test will depend on the problem under study.

Test statistic.

The test statistic is a number calculated from a statistical test of a hypothesis. It shows how closely your observed data match the distribution expected under the null hypothesis of that statistical test.The test statistic is used to calculate the *p*-value of your results, helping to decide whether to reject your null hypothesis.

Different Types of Test Statistics:

- t-statistic
- Z-statistic
- F-statistic

Z-TEST

- Formula to find the value of Z (z-test) Is:

$$Z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$$

- $\bar{x}$ = mean of sample
- μ_0 = mean of population
- σ = standard deviation of population
- n = no. of observations

t-test

Type	T - statistic	Degrees of freedom
One-sample t-test	$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$	df = n-1
Two-sample t-test	$t = \frac{\bar{X}_1 - \bar{X}_2}{s_{X_1X_2} \cdot \sqrt{\frac{1}{n1}+\frac{1}{n2}}}$	df = $n_1 + n_2 - 2$
Paired t-test	$t = \frac{\bar{X}_D - \mu_0}{s_D/\sqrt{n}}$	df = n-1

LeadQuizzes

p-values.

A p value is used in hypothesis testing to help you support or reject the null hypothesis. The p value is the evidence **against** a null hypothesis.

OR

The probability of happening of the event of null hypothesis is than the significance level, it has probably occurred by chance and this result is called statistical significance. This probability of happening of null hypothesis is calledp value..

P values gives us an idea of how strongly the data contradicts or supports the hypothesis. P values always helps us to access significance of any level.The smaller the p-value, the stronger the evidence that you should reject the null hypothesis.

P-Value

Now it is important to know how to find the p-value.

1. First we find the test statistic(t or Z test)
2. Then, with help of the cumulative distribution function (cdf) of the distribution. We find the corresponding cdf of the calculated t test or z test from their respective cdf table.
3. For a left tailed test, p value = cdf(x), suppose x is the calculated test statistics. (x is -ve)
4. For a right tailed test, p-value = 1 - cdf(x) (x is +ve)
5. For a two tailed test, p- value = 2*cdf(x), (x is-ve)

P-value Table

The P-value table shows the hypothesis interpretations:

P-value	Decision
P-value > 0.05	The result is not statistically significant and hence don’t reject the null hypothesis.
P-value < 0.05	The result is statistically significant. Generally, reject the null hypothesis in favour of the alternative hypothesis.
P-value < 0.01	The result is highly statistically significant, and thus rejects the null hypothesis in favour of the alternative hypothesis.

P-Value Example

Question: A statistician wants to test the hypothesis H0: $\mu = 120$ using the alternative hypothesis Hα: $\mu > 120$ and assuming that $\alpha = 0.05$. For that, he took the sample values as $n = 40$, $\sigma = 32.17$ and $\bar{x} = 105.37$. Determine the conclusion for this hypothesis?

Solution:

$$Z \text{ Test} = (\bar{x} - \mu) / (\sigma / \sqrt{n})$$

$$Z \text{ test} = (105.37 - 120) / (32.17/\sqrt{40})$$

$$= (105.37 - 120)/5.0865$$

$$= -14.63/5.0865$$

$$= -2.8762$$

Using the Z-Score table, we can find the value of $P(z > -2.8762)$

From the table, we get

$P(z > -2.8762) = 0.0021$

Therefore,

If $P(z > -2.8762) = 1 - 0.0021 = 0.9979$

P- value $= 0.9979 > 0.05$

Therefore, from the conclusion, if $p > 0.05$, the null hypothesis is accepted or fails to reject.

Hence, the conclusion is “fails to reject H_0.” Null hypothesis is true.

Example 1:

Question: In the National Academy of Archery, the head coach intends to improve the performance of the archers ahead of an upcoming competition. What do you think is a good way to improve the performance of the archers?

He proposed and implemented the idea that breathing exercises and meditation before the competition could help. The statistics before and after experiments are below:

	Before Experiment	After Experiment
Years	10	1
Mean	74	78
Std Dev	8	5
Observations	>1000	60

Interesting. The results favor the assumption that the overall score of the archers improved. But the coach wants to make sure that these results are because of the improved ability of the archers and not by luck or chance. So what do you think we should do?

Solution: Population Mean = 74, Population Standard Deviation = 8 ,Sample Mean = 78, Sample Size = 60

According to the problem above, there can be *two possible conditions*:

1. The after-experiment results are a matter of luck, i.e. mean before and after experiment are similar. This will be our "Null Hypothesis" .

$$H_0: \mu = \bar{X}$$

2. The after-experiment results are indeed very different from the pre-experiment ones. This will be our "Alternate Hypothesis".

$$H_0: \mu \neq \bar{X}$$

$$Z\ Test = (\bar{x} - \mu) / (\sigma / \sqrt{n})$$

$$= (78-74)/(8/\sqrt{60}) = 4/(1.03) = 3.88$$

Now we refer to the Z-table and find the p-value:

If we look up the Z-table for -3.88, we get a value of ~0.0005 (two-tailed test, so we have to multiply it by two).

p-value =2 * 0.0005

p-value = 0.001

We were not given any value for alpha, therefore we can consider alpha = 0.05. According to our understanding, if the *likeliness of obtaining the sample (p-value)* result is less than the alpha value, we consider the sample results obtained as significantly different.

Therefore, it is convenient to say that the increase in the performance of the archers in the sample population is not the result of luck.