Hypothesis testing is done for the following:

1. Hypothesis testing for a proportion.
2. HT for a mean when we know the sigma.
3. HT for a mean when we don't know the sigma.

Hypothesis:

This is some sort of claim. It is an educated guess. Hypothesis testing is testing a claim whether it is right or wrong, or whether the claim is valid.

What is a claim?

This is some sort of a statement. E.g. most people get their jobs through networking. Now, first we have to decide what we are dealing with here, mean or proportion in the statement? What does most mean here? Here, most, means more than 50%. That means "most" in the statement is referring to proportion. Then it would be:

$P > 0.5$

So, we have to decide always what the statement is telling us about, proportion or mean? It always proportion when there is most or percentage or less than some percentage. And "mean" means when the statement has average of something. E.g. the average payload of trucks on the 99 (freeway) is 18,00 lbs. We are dealing with mean here. We are talking about the whole population using a sample.

$\mu = 18{,}000$

Rare Event Rule:

We make assumptions out of this claim, and we are going to test the probability of these assumptions being true. If the probability of an assumption occurring is "very small", then the assumption is probably incorrect. We say here "probably", because we can never be 100% sure about an assumption to be wrong or correct. Statistics always talk about reasonably sure. That's where our confidence level comes.

Example:

Gender selection-

Claim- If you use drug "a", there is at least 80% chance of having a girl. (is this a proportion or average?)

Sample- 100 couples (they will be studied), according to our claim, 80 couples will have girls.

The assumption- The drug doesn't work. Let's say, 50% will be girls and 50% will be boys. For the following cases:

a) 52/100 had girls, this probability is not that different from the assumption that 50% will be girls.
b) 97/100 had girls, this probability is significantly different from the assumption that 50% will be girls which in fact disapproves our assumption that the drug doesn't work. If the probability of this action is rare considering the statement, then the statement is wrong that means it approves our claim that If you use drug "a", there is at least 80% chance of having a girl.

Hypothesis testing is a form of statistical inference that uses data from a sample to draw conclusions about a population parameter or a population probability distribution. First, a tentative assumption is made about the parameter or distribution. This assumption is called the null hypothesis and is denoted by H_0. An alternative hypothesis (denoted H_a), which is the opposite of what is stated in the null hypothesis, is then defined. The hypothesis-testing procedure involves using sample data to determine whether or not H_0 can be rejected. If H_0 is rejected, the statistical conclusion is that the alternative hypothesis H_a is true.

In hypothesis testing, we always try to prove a claim incorrect. For example, we can never prove someone innocent, what we do we prove them guilty or not guilty. In hypothesis testing, either we reject a claim, or we fail to reject a claim.

Types of Hypotheses:

Null hypothesis, H_0 (statement of equality)– This is always going to be a statement that reflects some population parameter (mean (μ), proportion (p)) is equal to a value. Null hypothesis will always have equal value. Example: $H_0 : \mu = 5$, $H_0 : p = 0.5$

How to test hypothesis?

We start by assuming that the null hypothesis is true. Then by evidence, we reach the following conclusions:

- Reject the null hypothesis (e.g. I have enough evidence to prove null hypothesis (H_0) wrong)
- Fail to reject the null hypothesis (e.g. I don't have enough evidence to prove the null hypothesis (H_0) wrong).

We can never accept H_0, either we have evidence to reject it or we don't have enough evidence to reject it that means fail to reject (H_0).

Now let's talk about alternative hypothesis H_a, how these things interplay with each other?

Alternative hypothesis, H_a (alternative hypothesis will never have equal sign in it)- States that the parameter (mean, proportion) has a value different than null hypothesis. (e.g. >, <, ≠).

Ex: $H_a : p < 0.5$, $H_a : p > 0.5$, $H_a : p \neq 0.5$

These two hypotheses work together. We are not proving anything directly in hypothesis testing rather we do it indirectly. Null and alternative hypotheses are opposite statements of each other. That means if we reject null hypothesis, we indirectly accept alternative hypothesis. We can never prove null hypothesis right/true, we can either reject it or fail to reject it. If we fail to reject null hypothesis that means we fail to accept the alternative hypothesis.

Note: If we want to prove a claim true/right, can we state it as null hypothesis? The answer is No. If we want to prove a claim/statement right/true, we must state it as an alternative hypothesis. Because alternative hypothesis is the only one, we can accept, and we can do so by proving null hypothesis wrong. If we want to support/prove a statement/claim, we must state it as an alternative hypothesis not null hypothesis.

Example: Suppose we want to prove that our previous fertility drug works (this is our alternative hypothesis because this is the claim/statement we want to prove right). We should state the null hypothesis of this claim and by only rejecting the null hypothesis we can say that the fertility drug works (which is our alternative hypothesis).

H_0: $p = 0.5$,

H_a: $p > 0.5$

If we fail to reject the null hypothesis, the result becomes inconclusive. That means we do not have enough evidence to reject the null hypothesis, it doesn't mean the null hypothesis is right or the alternative hypothesis is wrong, it means we don't have enough evidence to say anything.

How to identify null and alternative hypotheses?

First, we must state our original claim symbolically. Then state the opposite of the original claim as well. But this only can be done in research and project works. If the claim is given to you (as class work or exercise) you can't change that. When the claims are given to you, they can either be worded as null or alternative statement.

The statement/claim which has equality in it will be the null hypothesis. In the next example, first we will find out the statement is talking about proportion or mean, then we will state the original and opposite claim symbolically and finally we will decide which one is null and alternative.

Example: The mean of fluid is **at least** 12 oz in a can.

Is it talking about proportion or mean? The answer is mean.

Now, let's write the original and opposite claim symbolically.

Original claim- $\mu \geq 12$, Null hypothesis, H_0: $\mu = 12$ [here, we did a slight restatement. We completely omit the other greater or lesser sign where we have the equal]

Opposite claim- $\mu < 12$, Alternative hypothesis, H_a: $\mu < 12$ [in this case, we leave it as it was]

Note: These statements are about population. What we are doing, we are taking samples, using that as evidence to test a claim about the population. Why do we want to test a claim about a sample? We have all the information about the sample, we don't need to make a claim, we have everything there, but what we don't have is the information about the population. So, what do we do, we use this evidence to confirm or reject some claim about a population. We are basically testing claims about a population. These claims are all based on population parameters which are μ, p, σ not about any sample parameters.

Example: The mean of fluid is **more than** 12 oz in a can.

Now, let's write the original and opposite claim symbolically.

Original claim- $\mu > 12$, Alternative hypothesis, H_a: $\mu > 12$ [in this case, we leave it as it was]

Opposite claim- $\mu \leq 12$, Null hypothesis, H_0: $\mu = 12$ [here, we did a slight restatement. We completely omit the other greater or lesser sign where we have the equal]

Reminder: To prove a statement true, it should always be stated as an alternative hypothesis, H_a.

Example: The proportion of male CEO is greater than 0.5. (most always means more than 50% or 0.5, so if you ever see most it means proportion)

Original claim- $p > 0.5$, Alternative hypothesis, H_a: $p > 0.5$ [in this case, we leave it as it was]

Opposite claim- $p \leq 0.5$, Null hypothesis, H_0: $p = 0.5$ [here, we did a slight restatement. We completely omit the other greater or lesser sign where we have the equal]

Example: The mean weight of babies is at most 8.9 lbs.

Original claim- $\mu \leq 8.9$, Null hypothesis, H_0: $\mu = 8.9$ [here, we did a slight restatement. We completely omit the other greater or lesser sign where we have the equal]

Opposite claim- $\mu > 8.9$, Alternative hypothesis, H_a: $\mu > 8.9$ [in this case, we leave it as it was]

Example: The mean IQ score is 100.

Original claim- $\mu = 100$, Null hypothesis, H_0: $\mu = 100$ [here, we did a slight restatement. We completely omit the other greater or lesser sign where we have the equal]

Opposite claim- $\mu \neq 100$, Alternative hypothesis, H_a: $\mu \neq 100$ [in this case, we leave it as it was]

Test Statistics

Test statistics are what we use to test the hypothesis.

Proportion, p:

Z test. It is a typical Z- score. This is not the $\alpha/2$ critical value. For proportion, we always have Z-statistic only.

$$Z = \frac{\hat{p}-p}{\sqrt{\frac{pq}{n}}}$$

Where,

$\hat{p}$ = sample measure

p = population parameter (which we will get from our claim)

$\hat{p}$, and n are the only things that will come from our sample.

Mean, μ:

We have two options for mean, μ. We use Z-test if we know population's standard deviation, but if we don't know population's standard deviation, we use sample's standard deviation. Why? It is all about standard deviation because Z-score is based on standard deviation, σ. The selection of Z or T test also depends on the sample size. If the sample size is over 30, we use Z-test whereas we use T-test for a small sample size which is less than 30.

$Z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$ $\qquad\qquad$ $T = \frac{\bar{X} - \mu}{\frac{s}{\sqrt{n}}}$

Example: Survey: A sample of 706 companies found that 61% of CEOs were male. **Claim:** Most CEOs are male.

We have to test the claim. For doing that the question comes first is: Is 61% different/ big enough to say that most are male? E.g. is 52% or 55% big enough to say that most CEOs are male? The second question comes- Is the sample 706 is well representative of the population? Is the sample size large enough to say that most CEOs are male?

Step 1: Identifying null and alternative hypothesis

Now, let's write null and alternative hypothesis from our claim. (**Claim:** Most CEOs are male.)

Original claim- $p > 0.5$, H_a: $p > 0.5$ (why didn't we wrote $p > 0.61$?- The 61% result we have from our survey, so the 61% is talking about 706 sample companies. But the claim we made that most CEOs are male is based on the result we got from the survey. We are claiming about population considering the survey result. 61% is our evidence here, we are going to use that to confirm or deny our claim). **Alternative hypothesis, in our case p > 0.5 is always going to tell whether we are going to have a right tail/ a left tail/ or a two tailed test.**

Opposite claim- $p \le 0.5$, H_0: $p = 0.5$

Step 2: Test Statistics

$Z = \frac{\hat{p} - p}{\sqrt{\frac{pq}{n}}}$

Where,

$\hat{p}$ = sample measure (here it means the sample proportion. For our example the sample proportion is 0.61).

p = population parameter (which we will get from our claim and the p always comes from the null hypothesis or H_0. Here, according to our claim $p = 0.5$)

$q = 1-p$,

$= 1-0.5$

$= 0.5$

And, n = 706 (the sample size)

See, we told earlier that $\hat{p}$, and n are the only things that will come from our sample.

Z = $\frac{\mathbf{0.61-0.50}}{\sqrt{\frac{(\mathbf{0.50X0.50})}{\mathbf{706}}}}$ **(always do the top operation first)**

= 5.84

Now, we have to find out that 5.84 is a usual or unusual value for a Z-score.

Note: Z-score and unusual values:

- z scores are one of the common measures of relative position; they describe the location of a value relative to the mean. Here, the mean always has a zero z score.
- A z score of 1 indicates that a data value is one standard deviation above the mean, while minus 2 suggests two standard deviations below the mean.
- The ordinary, or majority, of values in any distribution lie within the z score of minus 2 to plus 2. Any values beyond this range are considered unusual, or outliers, and are considered far away from the other data values. Outliers may indicate variabilities in measurement or experimental errors.
- For example, a student's height has a plus 3.3 z score, or 3.3 standard deviations away from the class average, indicating that she is unusually tall for her class.

If Z-score is rare enough it means our null hypothesis or H_0 is wrong. If our null hypothesis is wrong, it means our H_a is right and that means we just proved our claim. Now, the problem is we don't know what rare enough means. We don't know if our Z-score, 5.84 is rare enough to prove our claim. Now, we will talk about the significance level because that will help us to understand what is rare enough. (The significance level is that we are 95% confident that the significance level is 0.05. So, basically, we are going to say that how certain we are about our hypothesis testing.

How to make a decision based on evidence?

Significance level:

In order to a make a decision, we need a significance level (α). Significance level is always calculated as (1-C), here, C means confidence level. The common significance level for 90% confidence level is 0.10, for 95% it is 0.05, and for 99% confidence level the significance level, α, is 0.01. The thing is do these always have to be like these. The answer is NO. The significance level can be anything based on confidence level, but the mentioned are the most common one.

Critical values:

The critical value is going to separate the rejection region from the failed to reject regions. Why do we need critical value? Because we are trying to reject our null hypothesis in order to accept the alternative hypothesis. If the test statistics, in the above example case the Z-score falls in the rejection region, we reject our null hypothesis, and if doesn't fall in the rejection region, we fail to reject the null hypothesis.

That's the decision we are going to make.

Critical values for different levels of significance:

Level of Significance (α)	0.10	0.05	0.01	0.005	0.002
z_α for 1-Tailed Tests	−1.28 or 1.28	−1.645 or 1.645	−2.33 or 2.33	−2.58 or 2.58	−2.88 or 2.88
z_α for 2-Tailed Tests	−1.645 and 1.645	−1.96 and 1.96	−2.58 and 2.58	−2.81 and 2.81	−3.08 and 3.08

Rejection Region:

If our test statistic falls within this rejection region area, we reject our null hypothesis. If our test statistic falls in the fail to the rejection region that mean we don't have enough evidence to reject the null hypothesis.

Throughout the calculations, we have two Z-values. One we get from the test statistic, and the other is from the critical value.

Types of tailed hypothesis tests:

There are three basic types of 'tails' that hypothesis tests can have:

- One-tailed test
  - Right-tailed test: where the alternative hypothesis includes a '>' symbol. The rejection region is in the right tail and the critical value will be always positive.
  - Left-tailed test: where the alternative hypothesis includes a '<' symbol. The rejection region is in the left tail and the critical value will always be negative.
- Two-tailed test: where the alternative hypothesis includes a ≠. The rejection regions are in both tails with positive and negative same values.

One-tailed hypothesis tests

A test of hypothesis where the area of rejection is only in one direction. In other words, when change is expected to have occurred in one direction, i.e expecting output either increase or to decrease.

If the level of significance is 0.05, a one-tail test allots the entire alpha (α) in the one direction to test the statistical significance. Since the statistical significance in the one direction of interest, it is also known as a directional hypothesis.

Reject the null hypothesis; If the test statistic falls in the critical region, that means the test statistic has a greater value than the critical value (for the right-tailed test) and the test statistic has a lesser value than the critical value (for the left tailed test).

Generally, one-tailed tests are more powerful than two-tailed tests; because of that, one-tailed tests are preferred.

The basic disadvantage of a one-tailed test is it considers effects in one direction only. There is a chance that an important effect may be missing in another direction. For example, a new material used in the production and checking whether the yield improved over the existing material. There is a possibility that new material may give less yield than the current material.

One tailed tests are further divided into:

Right-tailed test (Right tailed test is also called the upper tail test. A hypothesis test is performed if the population parameter is suspected to be greater than the assumed parameter of the null hypothesis.)

Left-tailed test (Left-tailed test is also known as a lower tail test. A hypothesis test is performed if the population parameter is suspected to be less than the assumed parameter of the null hypothesis.)

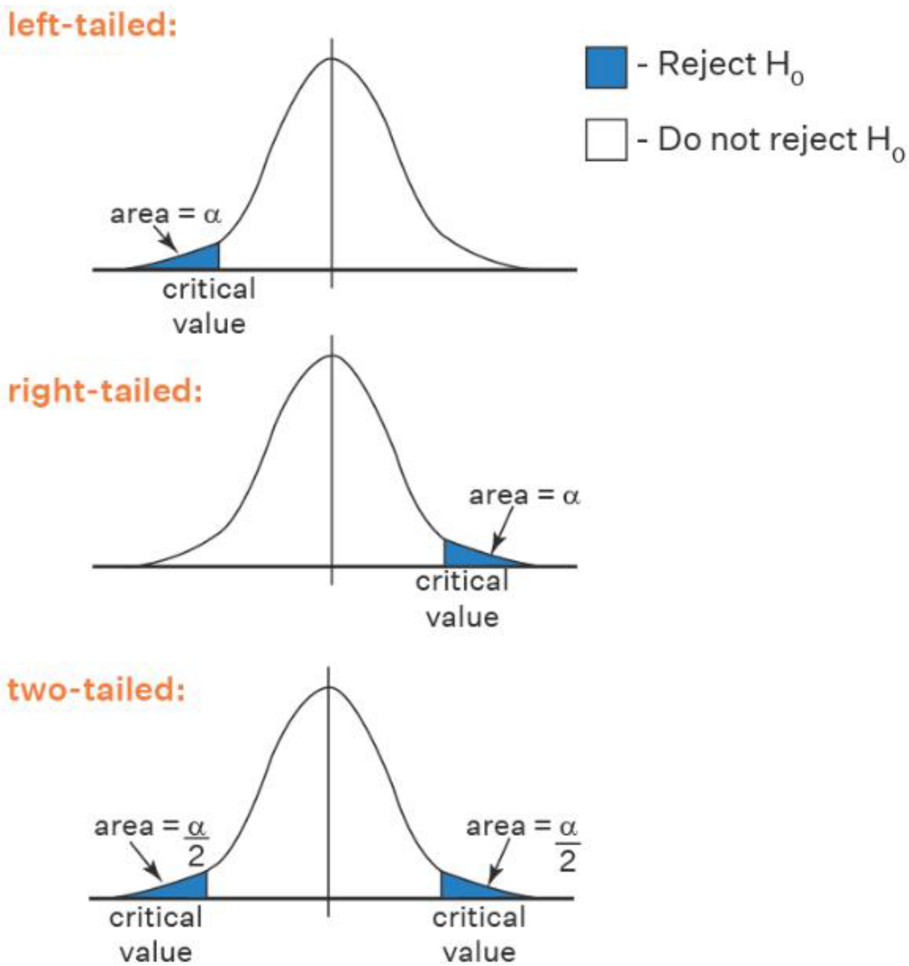
Two-tailed hypothesis tests

A test of hypothesis where the area of rejection is on both sides of the sampling distribution.

If level of significance is 0.05 of a two-tailed test, it distributes the alpha (α) into two equal parts ($\alpha/2$ & $\alpha/2$) on both sides to test the statistical significance.

Rejection Region for Null Hypothesis:

Let's say here the significance level,α is the common one which is 0.05. Considering the direction of the test, significance value and the critical value given above, we decide our reject and fail to reject region.



Now, let's get back to our example. The Z test statistic value we have is 5.84, the H_a: $p > 0.50$, and let's say the significance level is 0.05 (considering the confidence level 95%). Our test is right tailed as we know that the alternative hypothesis includes a '>' symbol. So, the critical value we have here is 1.645 (because α is 0.05 and the test is one tailed test which is in right direction). Our test statistic falls in the rejection region because the critical value starts from 1.645 and our test statistic value is 5.84.

P (probability)-value:

This probability value is associated with the test statistic.

There are two ways to do hypothesis testing. One- the traditional way, which we just did. We calculated our test statistic, we decided a significance level, and then considering the direction of the test we decided the critical value, and then finally decided the test statistic falls in the critical region or not. The p-value method does something backwards. In this method, we don't have to worry about the critical value. In this method, all we have our test statistic, and we are going to put that in our table/chart and we are going to find a value for our test statistic. Then we compare the value with the significance level. If it is smaller or equal to the significance level, then we know that it is rare enough to reject the null hypothesis. If it is greater than the significance value, then we fail to reject the null hypothesis.

Example: Find p- values.

Step 1- To determine whether it is right/ left or two tailed tests.

Step 2- To find the p- value.

Step 3- To compare the p- value with the significance level, α.

i) $\alpha = 0.05$, $H_a: p > 0.25$ (here, p refers to proportion. Alternative hypothesis doesn't talk about probability value). And the test statistic is $Z = 1.18$. After looking at the table we will get 0.8810 for the 1.18 z-score. Then p- value will be, 1- 0.8810, which is 0.1190. Now let's compare the p-value (0.1190) with the significance level (0.05). We can see that the p- value is greater than the significance level. If the p-value is greater than the significance value, then we fail to reject the null hypothesis.

Two (2) types

1. Type 1 Error: It refers when we reject null hypothesis while it is true (Wrongly reject H0).

E.g: H0: There is no difference between the effect of two new establish drugs.

Type 1 error occurs if we conclude, there is significance difference between the effects of two drugs when there isn't actually no differences.

Prob (Type 1 error) = Significance level = α

2. Type 2 Error: It refers when we accept null hypothesis while it is false.

E.g: H0: There is no significance difference in CKD prevalence between male and female.

Type 2 error occurs if we conclude that the prevalence of CKD in male and female is equal when there is actually differences.

Prob (Type 2 error) = β

Types of Errors

		Conclusion about null hypothesis from statistical test	
		Accept Null	Reject Null
Truth about null hypothesis in population	True	Correct	Type I error Observe difference when none exists
	False	Type II error Fail to observe difference when one exists	Correct

The top left quadrant means the ground truth is that the null hypothesis is true, yet our research concluded to *reject* the null hypothesis (*significant* effect from the treatment). We made an incorrect statistical decision. This is considered a Type 1 error.

The top right quadrant means the ground truth is that the alternative hypothesis is true and our research concluded to reject the null hypothesis (*significant* effect from the treatment). This is an ideal situation.

The bottom left quadrant means the ground truth is that the null hypothesis is *true* and our research concluded that there's *no* evidence to reject the null hypothesis. This is an ideal situation.

The bottom right quadrant means the ground truth is that the alternative hypothesis is true, yet our research concluded that there's *no* evidence to reject the null hypothesis. We made an incorrect statistical decision. This is considered a Type 2 error.