Hypothesis Testing

Basics and Tests

Hypothesis Testing

- Introduction to Hypothesis
- Type I and Type II Errors
- Test of Hypothesis

What is Hypothesis ?

What is Hypothesis ?

Example 1:
Pepsico India sells a soft drink in a can container having 250ml of the drink.

Since this information is on the label, we assume it to be true.

But is it really true ?

What is Hypothesis ?

Example 1:
Pepsico India sells a soft drink in a can container having 250ml of the drink.

As a customer:
We want the volume to be at least 250ml.
Quantity of Drink >= 250ml

As a manufacturer:
We want the volume of the drink to be exactly 250ml. No more, no less
Quantity of Drink = 250ml

What is Hypothesis ?

So we collected 100 Pepsi cans from all over India (to randomize the sample)

Then we measured the quantity of soft drink in each can in the sample and find the mean quantity for all 100 cans.

Using those sample means or measurements , we can TEST THE ASSUMPTION
(i.e. the STATUS QUO)

What is Hypothesis ?

Example 2:

An auto manufacturer had designed a new engine which claims to reduce the fuel consumption. It claims that the new engine makes more efficient use of fuel and performs better than the old engine which used to run at 30 miles per gallon.

Company now need to run some tests to look for statistical evidence to support the **claim** that the new engine offers better fuel efficiency.

What is Hypothesis ?

Company Claim:

Fuel Efficiency > 30

But it cold be well above or below 30 mpg.
The manufacturer is making a claim it wishes to test.

It is not testing an assumption
(i.e. Status Quo)

Notice
Assumption Vs. Claim

What is Hypothesis ?

When trying to formulate a statistical hypothesis, think about the following:

Are we testing an assumption (status quo) that already exists ? (Pepsi Can)
Or
Are we testing a claim (assertion) which may be true, if not then truth is something else ? (New Engine)

Say, if the manufacturer now tested the engine and fix it in a car and rolled it out in the market saying the car performs 35 mpg; and now we want to test it, we will be testing an assumption here and not a claim.

What is Hypothesis ?

Null Hypothesis:

The null hypothesis is the initial position. It is the status-quo position. It is the position that is rejected or fails to be rejected. It is the position that needs to be validated. It is the position that needs to be tested.

e.g.: Pepsi Can contains 250 ml of soft drink

Alternate Hypothesis:

The alternate hypothesis is the contrary position to NULL hypothesis. If there are statistically significant evidences that suggest that the alternate hypothesis is valid, then the NULL hypothesis is rejected.

e.g.: Pepsi Can does not contain 250 ml of soft drink (otherwise of NULL)

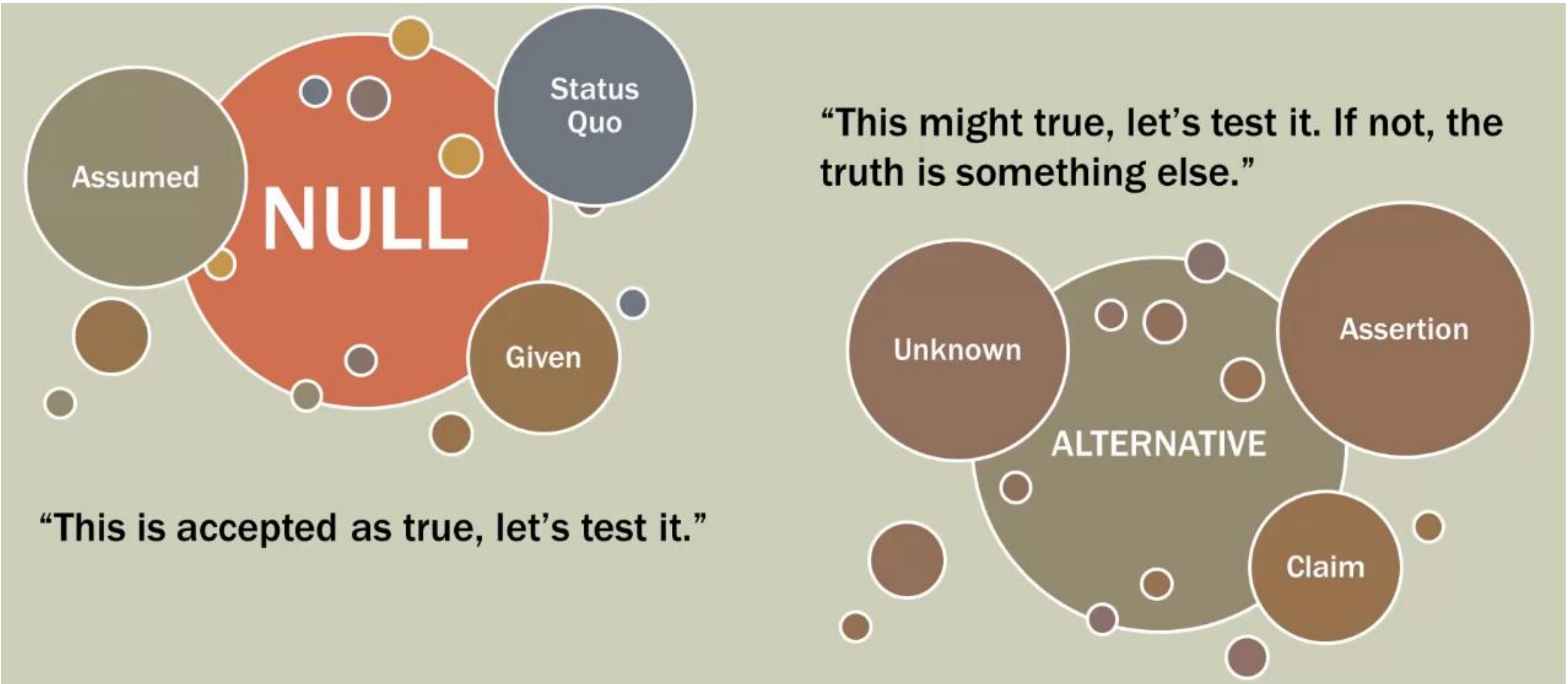
What is Hypothesis ?

By Definition, null and alternate hypothesis are opposite; mutually exclusive. They both cannot be true.

The null is either rejected or not. Only if the null is rejected can we move with the alternate hypothesis.

Null and alternate hypothesis depends on the question we are trying to ask. A person can start with either the null hypothesis or the alternate hypothesis, and then form the other as a complement to the first.

What is Hypothesis ?



What is Hypothesis ?

Hypothesis testing can be used to determine whether a statement about the value of a population parameter should or should not be rejected.

Null hypothesis: denoted by H_0 , is a tentative assumption about a population parameter.

Alternate hypothesis: denoted by H_a , is the opposite of what is stated in the null hypothesis.

e.g. Let's say a court room trial:

H_0 : The accused is innocent

H_a : The accused is guilty

What is Hypothesis ?

H_0	H_a
Assumption, status quo, nothing new	Rejection of an assumption
Assumed to be "true"; a given.	Rejection of an assumption or the given.
Negation of the research question	Research question to be "proven"
Always contains an equality $(=, \leq, \geq)$	Does not contain equality $(\neq, <, >)$

Based on the last property, we can derive the possible null/alternative pairs:

$$H_0:\ \mu \geq \mu_0 \qquad H_0:\ \mu \leq \mu_0 \qquad H_0:\ \mu = \mu_0$$

$$H_a:\ \mu < \mu_0 \qquad H_a:\ \mu > \mu_0 \qquad H_a:\ \mu \neq \mu_0$$

What is Hypothesis ?

Example:
The NY city provides one of the most comprehensive emergency medical services in the world. Operating in a multiple hospital system with approximately 200 mobile medical units, the service goal is to respond to medical emergencies with a mean time of 12 minutes or less.

The director of medical services wants to formulate a hypothesis test that could use a sample of emergency response times to determine whether or not the service goal of 12 minutes or less is being achieved.

What is Hypothesis ?

Hypothesis	Conclusion & Action
H_0 : $\mu <= 12$	The response time goal is met by Emergency Service team. No action required
H_a : $\mu > 12$	The response time goal is not met by Emergency Service team. Appropriate follow up actions are required.

where,
μ = mean response time for the population of medical emergency response unit.

What is Hypothesis ?

Few Important Guidelines:
All statistical conclusions are made in reference to the null hypothesis.

We either **reject** the null hypothesis or **fail to reject** the null hypothesis; we **do not accept** the null hypothesis.

If we **reject** the null hypothesis, then we conclude the data supports the alternative hypothesis.

However, if we **fail to reject** the null hypothesis, it does not mean we have proven the null hypothesis is “true”

What is Hypothesis ?

Example 1:

Is there anything we can assume to be true ?

Yes. 250 ml in container can be assumed to be true

Which hypothesis pair to be selected ?

$$H_0:\ \mu = \mu_0$$
$$H_a:\ \mu \neq \mu_0$$

$\mathbf{H_0}$ = 250ml

$\mathbf{H_a}$ ≠ 250 ml

If the cans are filled properly, then we **fail to reject** the null hypothesis.

We are not saying that we have proven the null.

Just our assumption has held up.

What is Hypothesis ?

Example 1:

According to the United States Department of Agriculture, in 2009 the average farm size in Texas was 2.3 sq. km. Due to large agriculture business in the past decade, the department wants to know if the current (2019) farm size is larger than it was in 2009.

Establish a null and alternative hypothesis.

Solution:

What is our assumption ?

We assume that there has been no change in the farm size since 2009.

This is our null hypothesis.

$$H_0: \mu \leq \mu_0$$
$$H_a: \mu > \mu_0$$

$\mathbf{H_0} \leq 2.3$ sq. km

$\mathbf{H_a} > 2.3$ sq. km

Type I and Type II Errors

Type I and Type II Errors

Since hypothesis tests are based on sample data, we must allow for the possibility of errors.

Type I error: False Positive

This error occurs when we reject the assumption (null hypothesis) when it should not have been rejected.

Type II error: False Negative

When we fail to reject the assumption (null hypothesis) when it should have been rejected.

Type I and Type II Errors

Example: NY Medical Emergency Unit
Our Hypothesis was:
$\mathbf{H_0}$ **:** $\mu <= 12$

Our Alternate Hypothesis was:
$\mathbf{H_a}$ **:** $\mu > 12$

	Population Condition	
Conclusion	**H_0 True** ($\mu <= 12$)	**H_0 False** (H_a True) ($\mu > 12$)
Accept $\mathbf{H_0}$	True Positive	False Negative **Type II Error**
Reject $\mathbf{H_0}$	False Positive **Type I Error**	True Negative

Type I and Type II Errors

Example:

Visualizing Type I and Type II Errors

Visualizing Type I and Type II Errors

Case:
Let's say we have population mean μ for a population under analysis.
The hypothesized mean for the population under analysis is μ_0.
Test if the hypothesized mean comes from the same population under analysis.

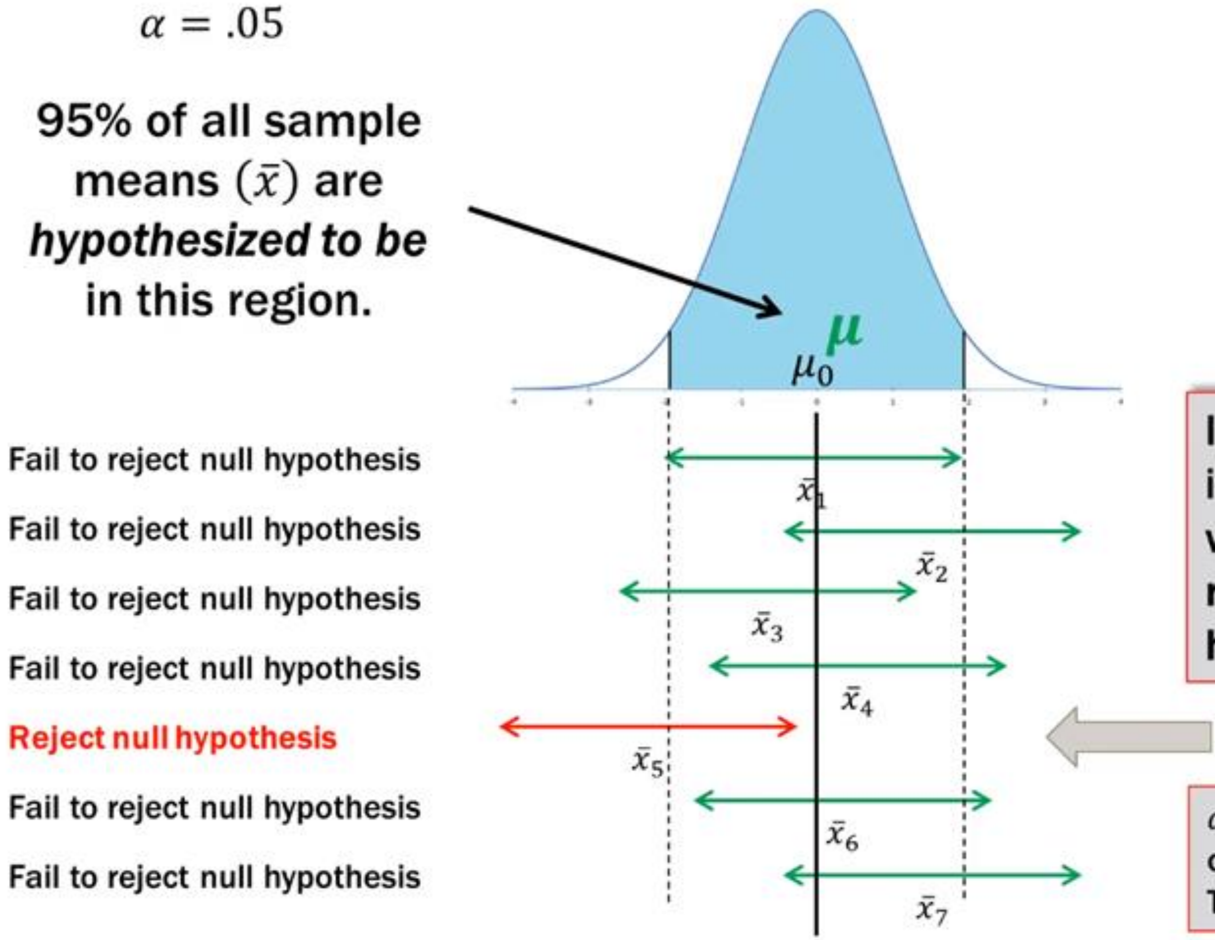
$H_0: \mu = \mu_0$ $H_a: \mu \neq \mu_0$

Visualizing Type I and Type II Errors

$\alpha = .05$

95% of all sample means ($\bar{x}$) are *hypothesized to be* in this region.

$$H_0\colon \mu = \mu_0$$
$$H_a\colon \mu \neq \mu_0$$



If we took a sample and it was by chance like $\bar{x}_5$, we would incorrectly reject the null hypothesis.

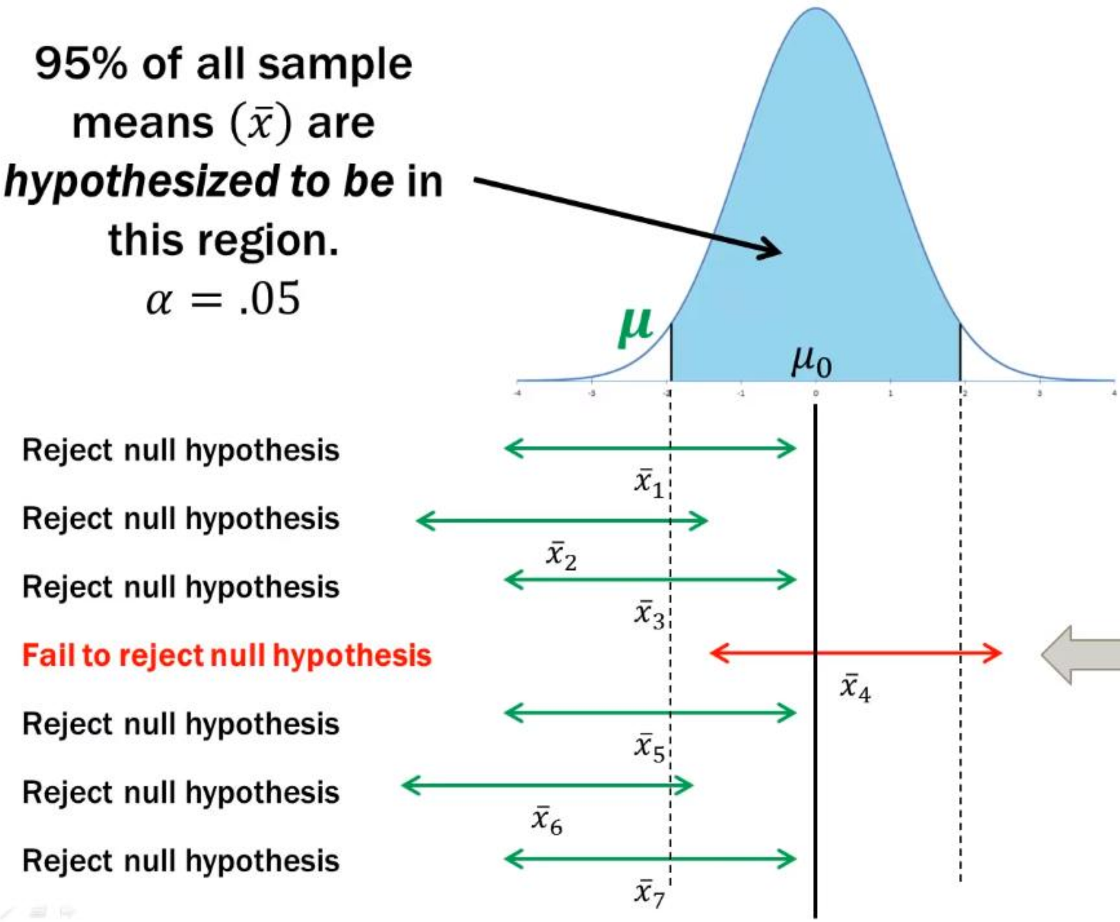
Type I Error

α is the "level of significance" or our tolerance for making a Type I error.

Visualizing Type I and Type II Errors

95% of all sample means ($\bar{x}$) are ***hypothesized to be*** in this region.

$\alpha = .05$



$$H_0: \mu = \mu_0$$

$$H_a: \mu \neq \mu_0$$

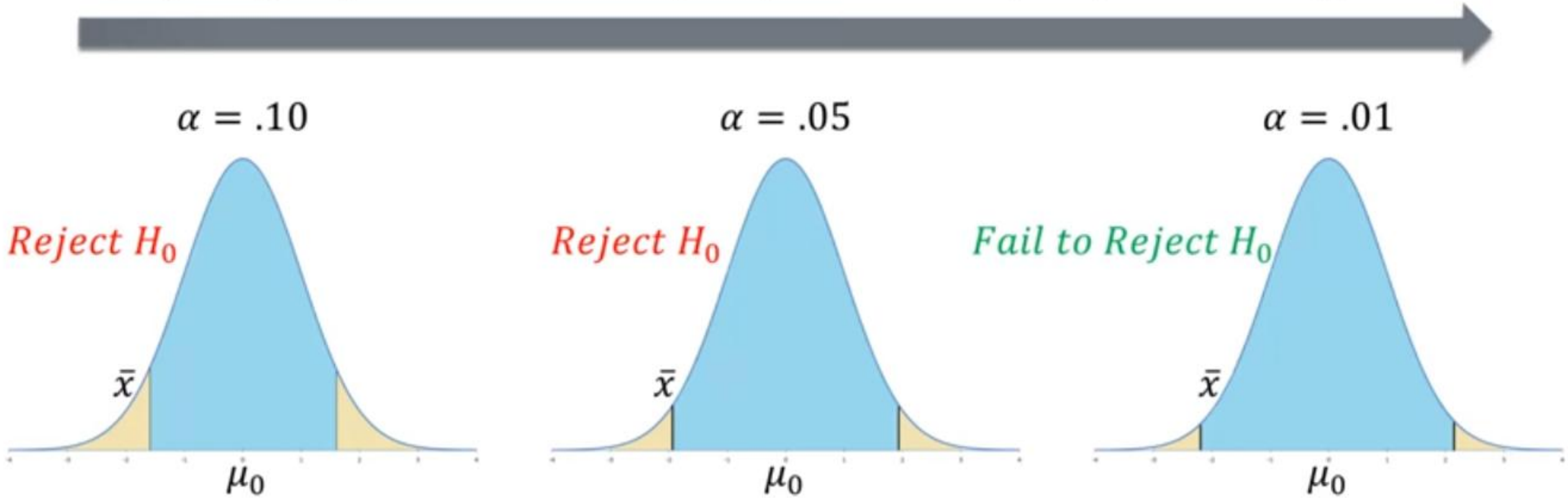
If we took a sample and it was by chance like $\bar{x}_4$, we would incorrectly "accept" the null hypothesis.

Type II Error

Beta (β) is the probability of committing a Type II error. The value of β varies with certain experimental factors.

Visualizing Type I and Type II Errors

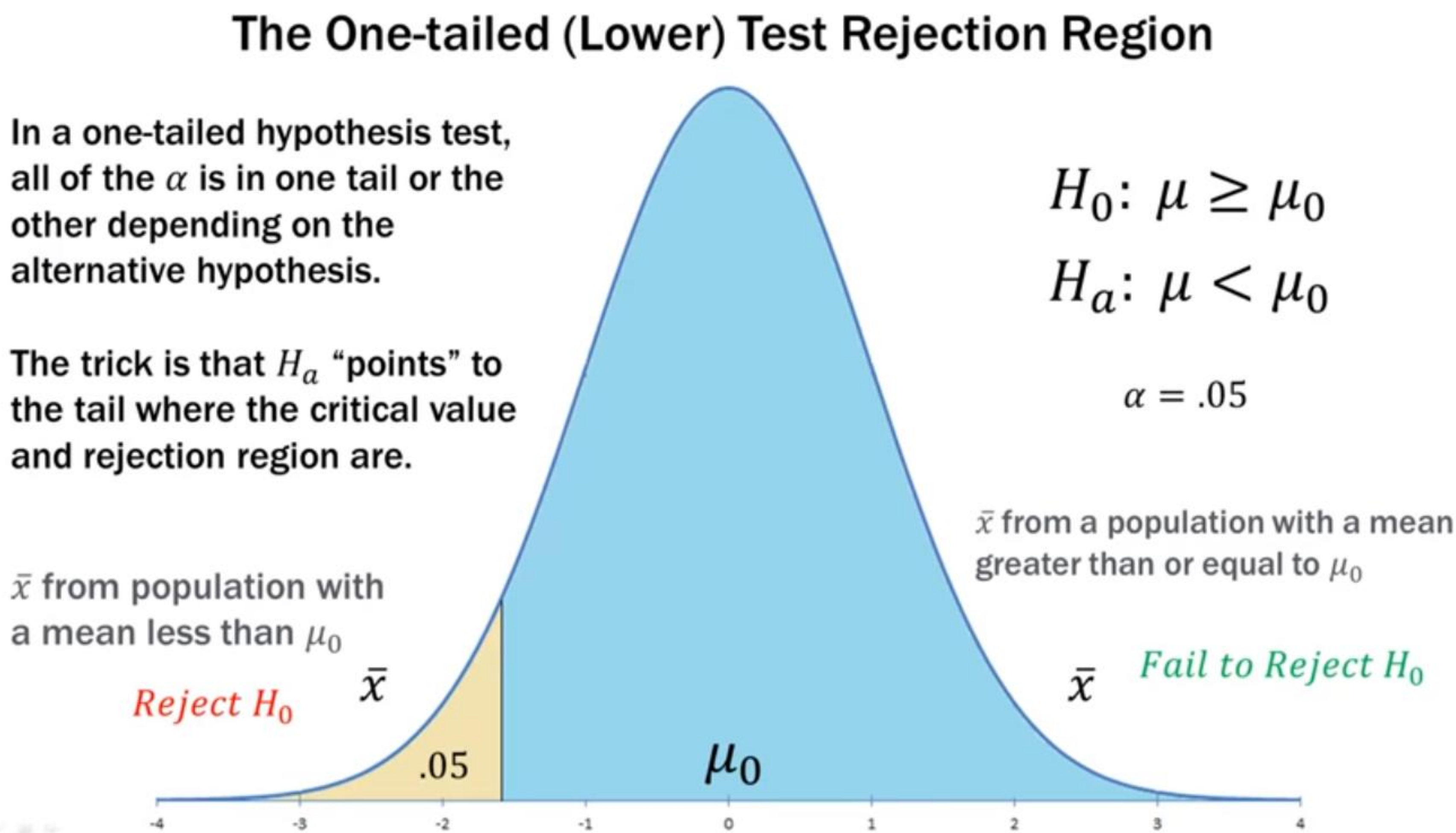
As α decreases so does the chance of Type I error. The critical value to reject the null hypothesis moves outward thus "capturing" more sample means. Kind of like moving the goal posts on a football pitch; the wider they are, more kicks go in.



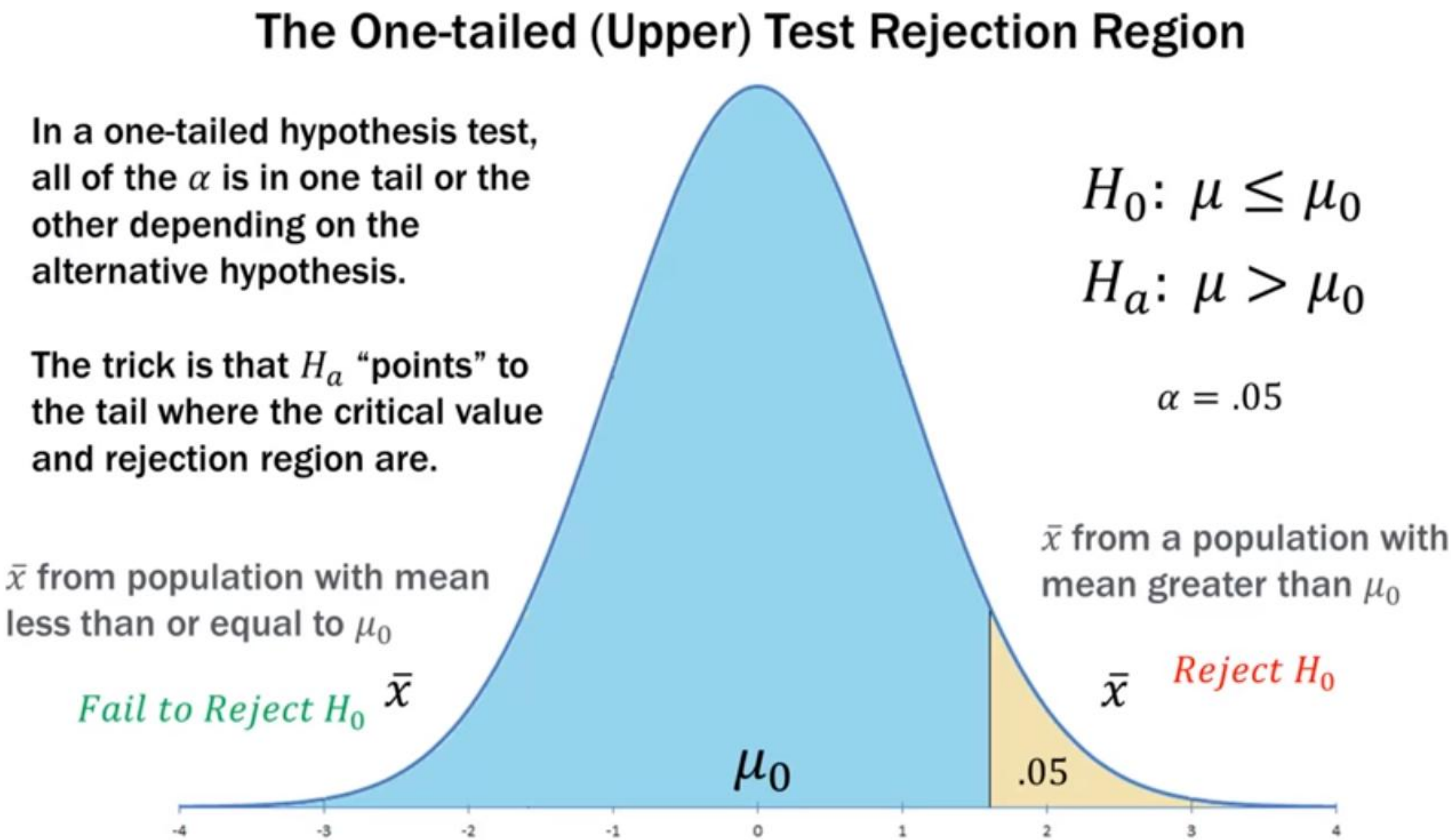
However the move outward of the critical values may also "capture" a mean from a different population off to the side. We would fail to reject the null when indeed we should. Thus the chance of Type II error increases as α decreases.

Visualizing One Tailed Test Rejection

One Tailed Test Rejection

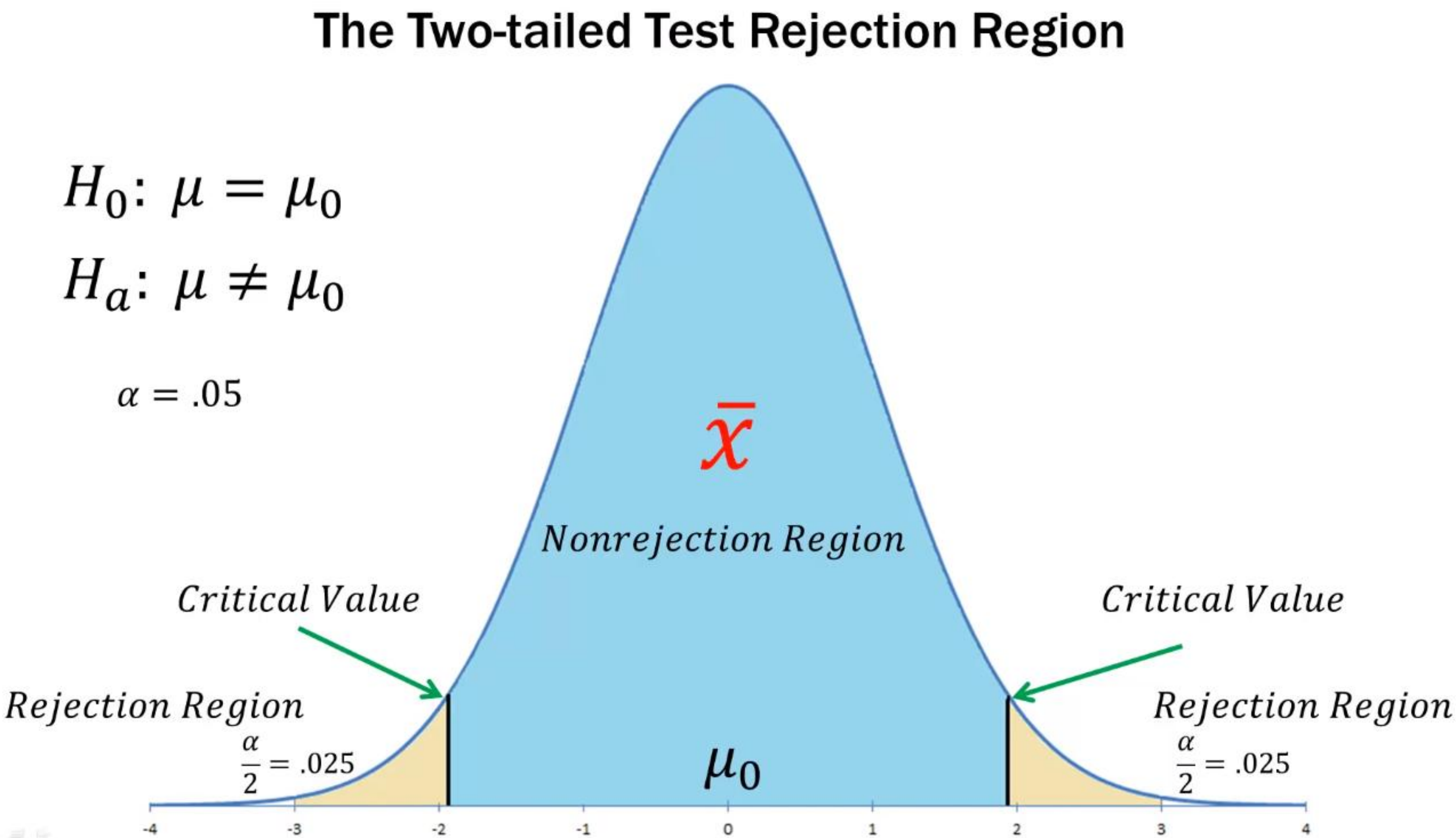


One Tailed Test Rejection



Visualizing Two Tailed Test Rejection

Two Tailed Test Rejection



Test of Hypothesis

Test of Hypothesis

Test of Hypothesis

- Large Sample Tests
  - Testing Population Mean (One Sample)
  - Testing Population Mean (Two Sample)
- Small Sample Test
  - t-test
    - One Sample
    - Two Sample
  - Paired t-test
  - Chi-square test
  - F - test

Test of Hypothesis

Large Sample Test

- Tests are based on large sample on which hypothesis is test
  (sample size n at least >30)

- Exact Sampling Distributions are not available.

- Test are based on C.L.T distribution.

Small Sample Test

- Tests are based on small sample on which hypothesis is test
- Exact Sampling Distributions are available.

The **exact sampling distribution** is the sampling distribution of a statistic (estimator, e.g. sample mean x bar) over all the possible sample drawn from a population (finite or infinite).

When to use z Vs t - distribution

When σ (sigma of population) is known -> use standard normal or z distribution

When σ (sigma of population) is unknown -> use t-distribution

Important:
Some books indicate the usage of z-distribution is acceptable any time when n >= 30, irrespective of sigma known or unknown.

However you can use the t-distribution anytime you do not know sigma, irrespective of n.

The reason behind this is, as the sample size increases, the z-distribution and t-distribution actually converges.

Large Sample Test

Testing Population Mean (μ) : One Sample

When to use ?

If you want to test whether or not a given sample is taken from a population whose mean value is specified (μ)

How to calculate statistics:

$$\boxed{z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}}}$$

Where,

$\bar{x}$ = Sample mean of ‘n’ observations.

σ = Population S.D. (known)

μ = Population mean

Testing Population Mean (μ) : One Sample

$$z = \frac{\bar{x} - \mu_0}{\underbrace{\frac{\sigma}{\sqrt{n}}}_{\text{Standard Error of the Mean (standard deviation of the sampling distribution)}}}$$

$$z = \frac{\bar{x} - \mu_0}{\sigma_{\bar{x}}}$$

$\bar{x}$ = sample mean

μ_o = hypothesized population mean

σ = population standard deviation (given)

n = sample size

Is this z-test value in the nonrejection region or the rejection region?

Hypothesis on One Sample z test

Here we test whether the given sample belongs to population having mean = μ_0

μ_0 =Specified value of mean

- H_0: Mean is equal to μ_0
- H_a: Mean is not equal to μ_0

Hypothesis on One Sample z test

Example

A report from 6 years ago indicated that the average gross salary for a business analyst was $69,873. Since this survey is now outdated, the Bureau of Labor Statistics wishes to test this figure against current salaries to see if the current salaries are statistically different from the old ones.

Based on other studies, we will assume $\sigma = \$13{,}985$.

For this study, the BLS will take a sample of 112 current salaries.

Hypothesis on One Sample z test

Solution

Step 1: Establish Hypothesis

$$H_0: \mu = \$69{,}873 \qquad H_a: \mu \neq \$69{,}873$$

Step 2: Determine Appropriate Statistical Test and Sampling Distribution

This will be a two-tailed test. Salaries could be higher OR lower.

Since σ is known, we will use the z-distribution.

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

Step 3: Specify the Type I error rate (significance level)

$$\alpha = .05$$

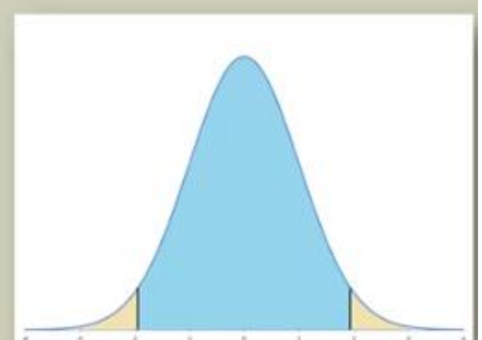
Step 4: State the decision rule

If $z > 1.96$, reject H_0

If $z < 1.96$, reject H_0

Step 5: Gather data

$$n = 112, \bar{x} = \$72{,}180$$



The mean here (x bar) looks higher. **But the question is: Is it high enough to be statistically significant ??**

Hypothesis on One Sample z test

Solution

The mean here (x bar) looks higher. **But the question is: Is it high enough to be statistically significant ??**

$\bar{x} = \$72{,}180$

$\mu_o = \$69{,}873$

$\sigma = \$13{,}985$

$n = 112$

$$z = \frac{\$72{,}180 - \$69{,}873}{\frac{\$13{,}985}{\sqrt{112}}}$$

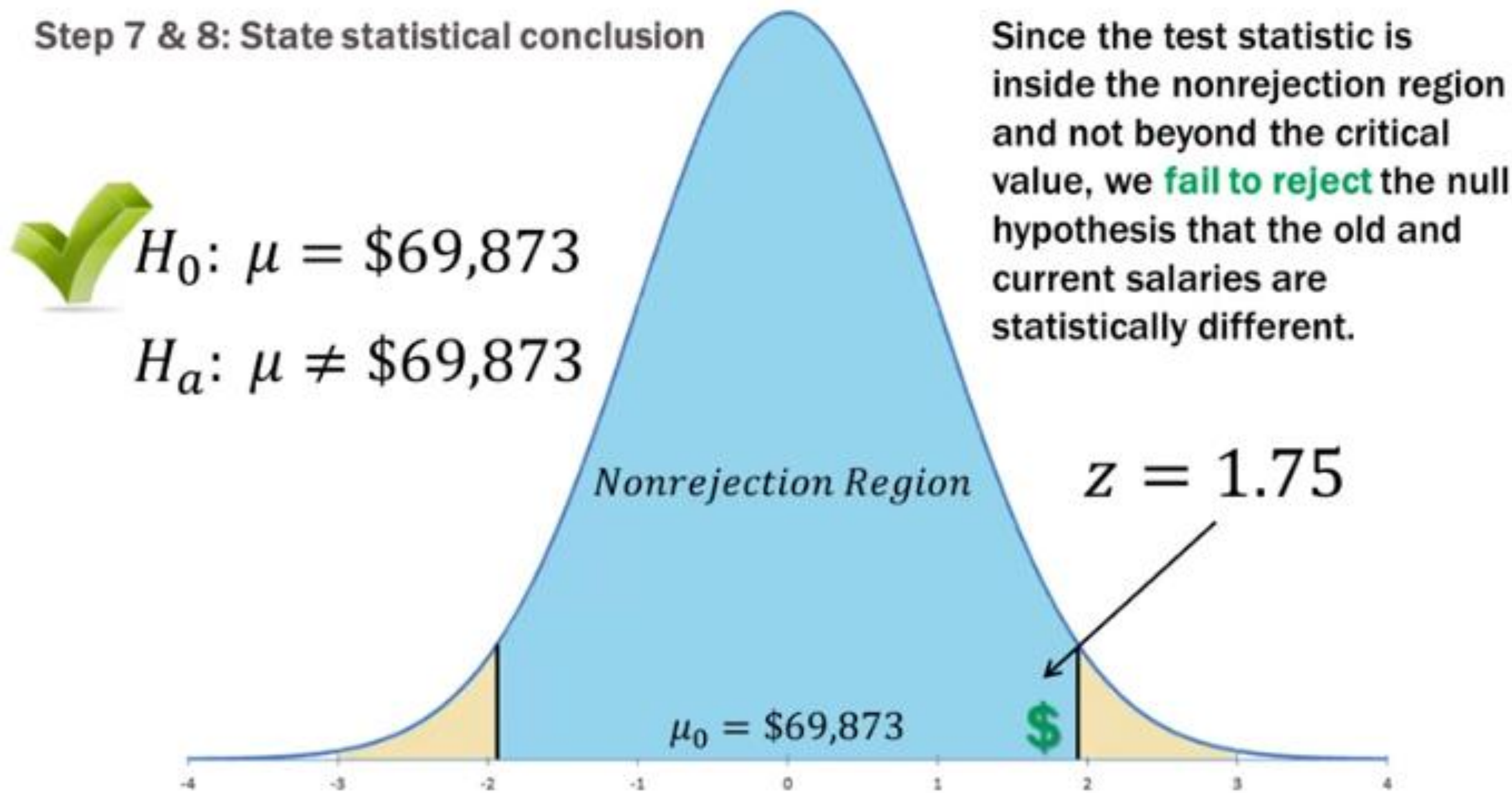
$$z = \frac{\bar{x} - \mu_0}{\frac{\sigma}{\sqrt{n}}}$$

$$z = 1.75$$

The z value from the z – table for alpha = 0.05 and hence Confidence Interval is '1.96' and our z critical value is '1.75' which is less. Hence it falls within the non-rejection region. **Hence, we fail to reject Null hypothesis that the old and current salaries are statistically different.**

Hypothesis on One Sample z test

Solution



Testing Population Mean (µ) : Two Sample

When to use ?

If you want to test whether if two population means are equal or unequal.

How to calculate statistics:

$$z = \frac{(\overline{x}_1 - \overline{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

Where,

$\overline{x}$ = Sample mean of 'n' observations.

σ1 = Population S.D. (known) of sample one

µ1 = Population mean of sample one

2 = Sample mean of 'n2' observations.

σ2 = Population S.D. (known) of sample two

µ2 = Population mean of sample one

Small Sample Test

Small Sample Test

In small sample test, the sampling distribution that the test statistic follow are majorly classified in three distributions:

- X^2 distribution (chi-square)
- t-distribution
- F-distribution

Small Sample Test: One Sample t-test

When to use ?

If you want to test whether a given sample is coming from a population or not, whose mean is specified value and population variance is unknown.

$$t = \frac{\bar{x} - \mu_0}{s / \sqrt{n}}$$

Where,

$\bar{x}$ = Sample mean of ‘n’ observations.

s = S.D. of sample of ‘n’ observations (sample).

μ_0 = Population mean.

Small Sample Test: One Sample t-test

$$t = \frac{\bar{x} - \mu_0}{\underbrace{\frac{s}{\sqrt{n}}}_{\text{Standard Error of the Mean (standard deviation of the sampling distribution)}}}$$

$$t = \frac{\bar{x} - \mu_0}{s_{\bar{x}}}$$

$\bar{x}$ = sample mean

μ_o = hypothesized population mean

s = sample standard deviation

n = sample size

Is this t-test value in the nonrejection region or the rejection region based on $df = n - 1$?

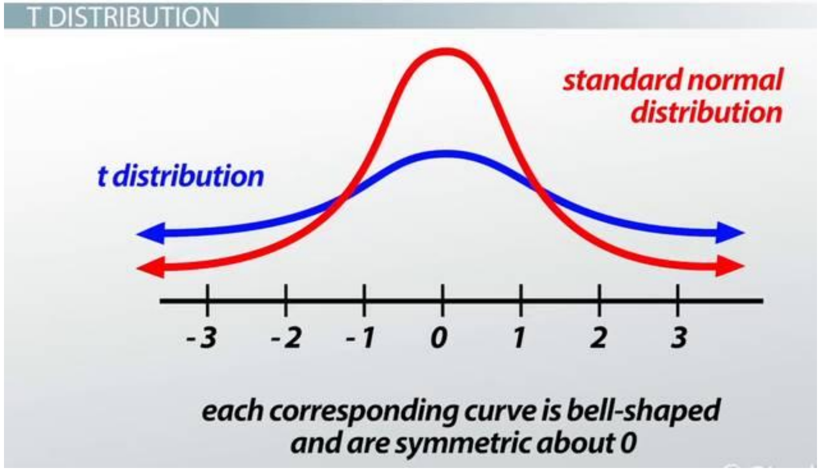
Hypothesis on : One Sample t test

Here we test whether the given sample belongs to population having mean = μ_0

μ_0 =Specified value of mean

- H_0: Mean is equal to μ_0
- H_a: Mean is not equal to μ_0

t – distribution properties



- A smaller sample size means more sampling error
- This sampling error due to small n means a higher probability of extreme sample means
- More probability in the tails means the center hump of the t-distribution must come downward.
- This process "squishes" the distribution slightly downward and outward thus taking the critical values along for the ride
- Given the same α and s, a smaller n will push the critical values further outward in the tail(s) due to the uncertainty associated with small n

Small Sample Test: One Sample t-test

Example:

A report from 6 years ago indicated that the average gross salary for a business analyst was $69,873. Since this survey is now outdated, the Bureau of Labor Statistics wishes to test this figure against current salaries to see if the current salaries are statistically different from the old ones.

Based on this sample, we found $s = \$14{,}985$. We do not know σ and will therefore have to estimate it using s.

For this study, the BLS will take a sample of 12 current salaries.

Small Sample Test: One Sample t-test

Solution:

Step 1: Establish Hypothesis

$$H_0\colon \mu = \$69{,}873 \qquad H_a\colon \mu \neq \$69{,}873$$

Step 2: Determine Appropriate Statistical Test and Sampling Distribution

This will be a two-tailed test. Salaries could be higher OR lower.

Since σ is unknown and n is small, we will use the t-distribution.

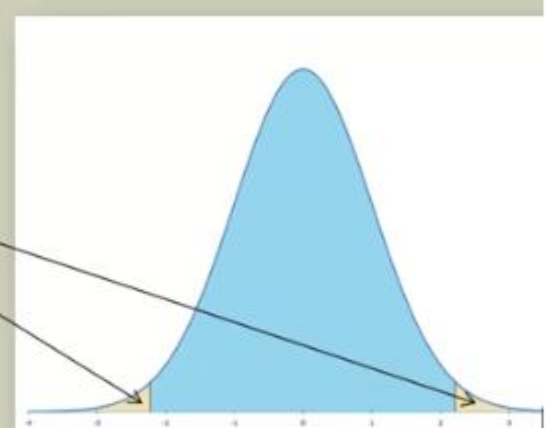
$$t = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}}$$

Step 3: Specify the Type I error rate (significance level)

$$\alpha = .05$$

Step 4: State the decision rule

$$\text{For } df = 11 \begin{cases} \text{If } t > 2.201, \text{ reject } H_0 \\ \text{If } t < -2.201, \text{ reject } H_0 \end{cases}$$



Step 5: Gather data

$$n = 12, \bar{x} = \$79{,}180$$

Small Sample Test: One Sample t-test

Solution:

Step 6: Calculate test statistic

$\bar{x} = \$79{,}180$

$\mu_o = \$69{,}873$

$s = \$14{,}985$

$n = 12$

$$t = \frac{\$79{,}180 - \$69{,}873}{\frac{\$14{,}985}{\sqrt{12}}}$$

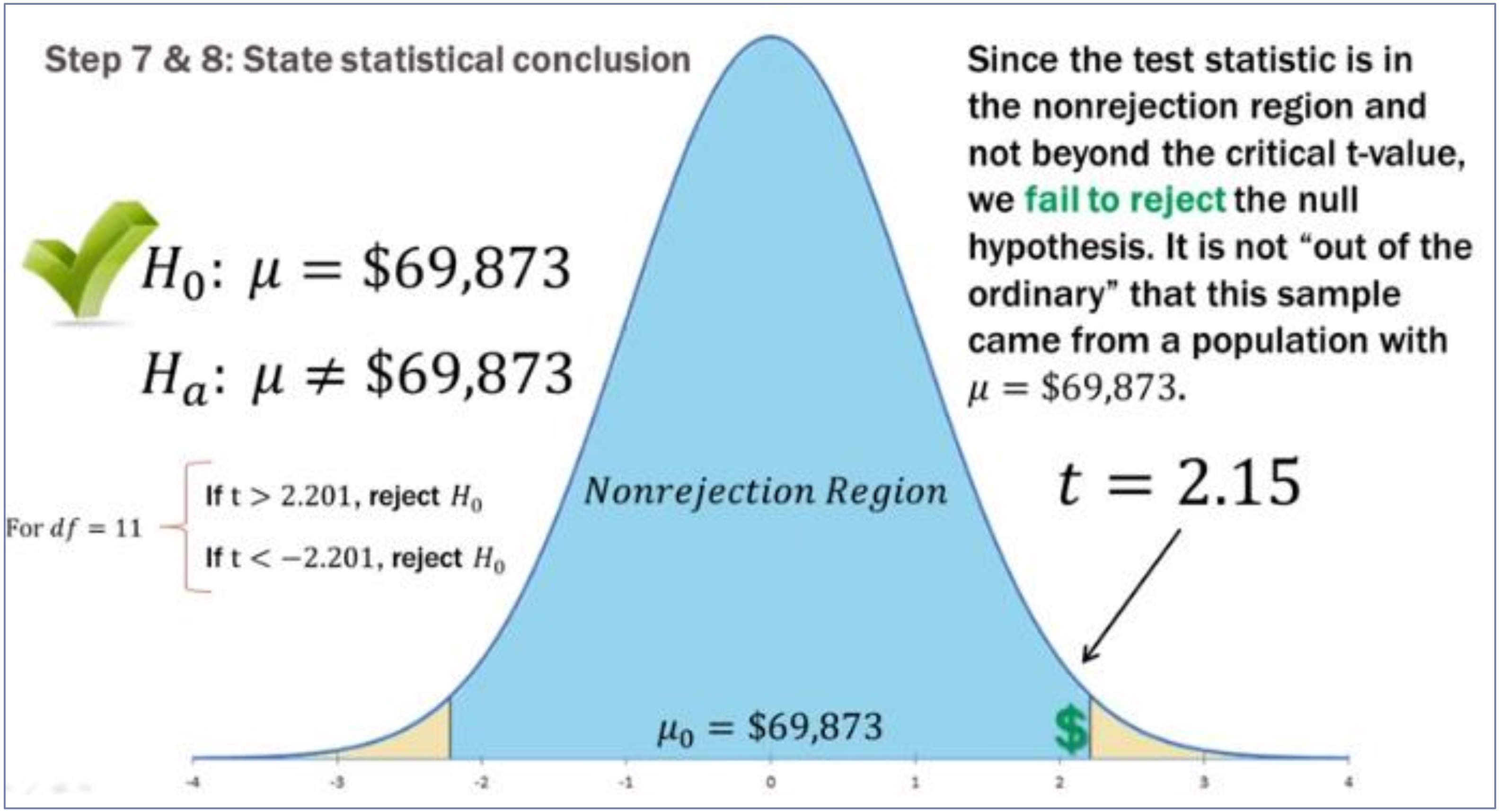
$$t = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}}$$

$$t = 2.15$$

Since the t value is ‘2.15’ which is lower than the t value of ‘2.201’ for alpha of 0.05, hence **we fail to reject the null hypothesis. Which means that this sample do come out of the sample with mean of $69,873.**

Small Sample Test: One Sample t-test

Solution:



Small Sample Test: Two Sample t-test

When to use ?

If you want to test whether given two samples belongs to same population or not. This can be tested by checking equality of means of two samples.

$$t = \frac{\bar{x}_1 - \bar{x}_2}{s_p * \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

Where,

$\bar{x}_1$ = Sample mean of n_1 observations

$\bar{x}_2$ = Sample mean of n_2 observations

S_p = Pooled S.D. of two samples

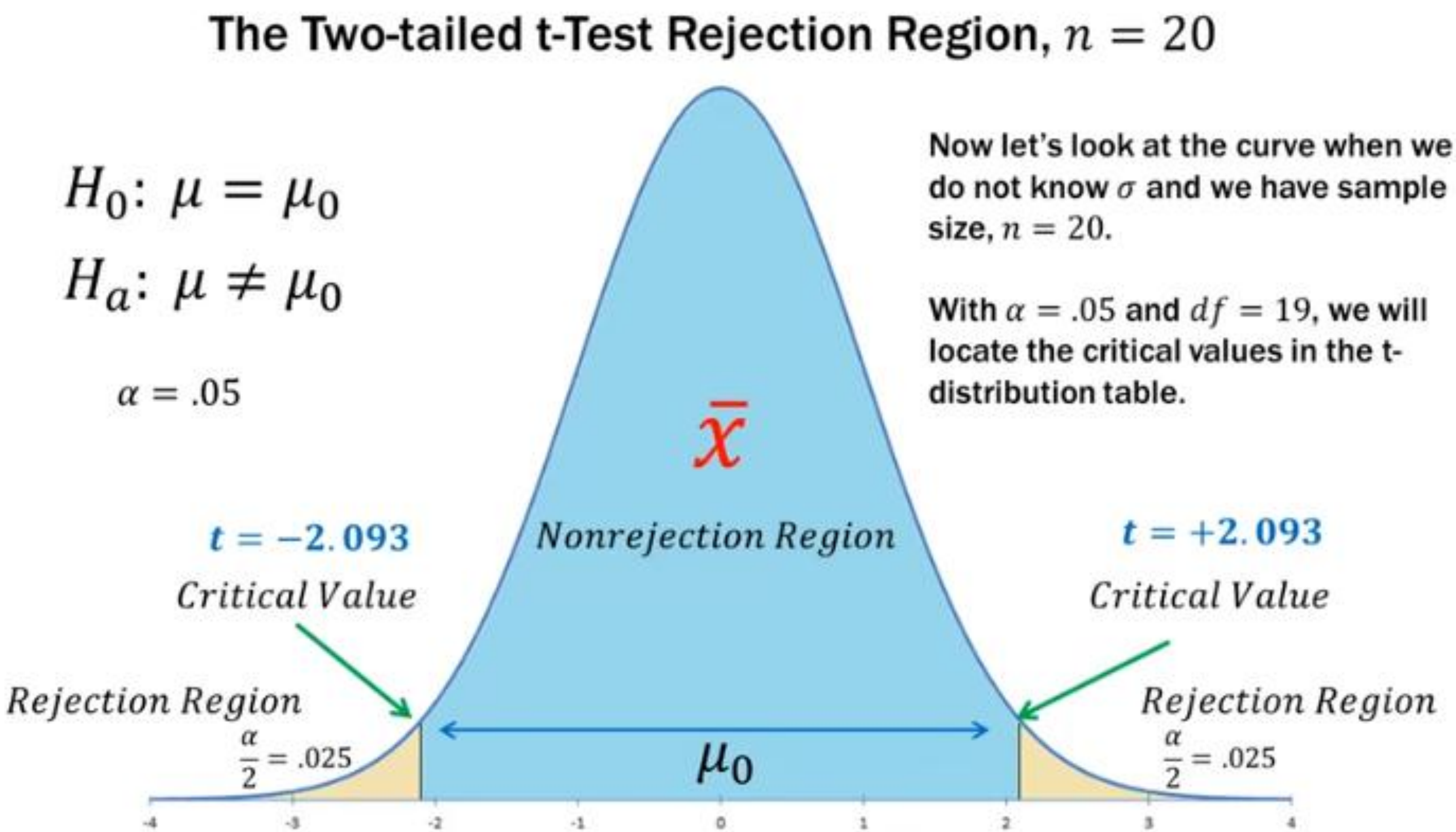
Hypothesis on : Two Sample t test

Here we test whether the given two sample belongs to same population.

- H_0: Mean of two samples is same.
- H_a: Mean of two samples is not same.

- H_0: Two samples belongs to same population.
- H_a: Two samples belongs to different population.

Hypothesis on : Two Sample t test



Hypothesis on : Two Sample t test

Looking up t-values

Now let's look at t-table when we do not know σ and we have sample size, $n = 20$.

With $\alpha = .05$ and $df = 19$, we will locate the critical values in the t-distribution table.

t Distribution

Degrees of freedom	α .005 (one tail) .01 (two tails)	.01 (one tail) .02 (two tails)	.025 (one tail) .05 (two tails)	.05 (one tail) .10 (two tails)	.10 (one tail) .20 (two tails)	.25 (one tail) .50 (two tails)
1	63.657	31.821	12.706	6.314	3.078	1.000
2	9.925	6.965	4.303	2.920	1.886	.816
3	5.841	4.541	3.182	2.353	1.638	765
4	4.604	3.747	2.776	2.132	1.533	741
5	4.032	3.365	2.571	2.015	1.476	727
6	3.707	3.143	2.447	1.943	1.440	.718
7	3.500	2.998	2.365	1.895	1.415	.711
8	3.355	2.896	2.306	1.860	1.397	.706
9	3.250	2.821	2.262	1.833	1.383	.703
10	3.169	2.764	2.228	1.812	1.372	.700
11	3.106	2.718	2.201	1.796	1.363	.697
12	3.054	2.681	2.179	1.782	1.356	.696
13	3.012	2.650	2.160	1.771	1.350	.694
14	2.977	2.625	2.145	1.761	1.345	.692
15	2.947	2.602	2.132	1.753	1.341	.691
16	2.921	2.584	2.120	1.746	1.337	.690
17	2.898	2.567	2.110	1.740	1.333	.689
18	2.878	2.552	2.101	1.734	1.330	.688
19	2.861	2.540	2.093	1.729	1.328	.688
20	2.845	2.528	2.086	1.725	1.325	.687
21	2.831	2.518	2.080	1.721	1.323	.686
22	2.819	2.508	2.074	1.717	1.321	.686
23	2.807	2.500	2.069	1.714	1.320	.685
24	2.797	2.492	2.064	1.711	1.318	.685
25	2.787	2.485	2.060	1.708	1.316	.684
26	2.779	2.479	2.056	1.706	1.315	.684
27	2.771	2.473	2.052	1.703	1.314	.684
28	2.763	2.467	2.048	1.701	1.313	.683
29	2.756	2.462	2.045	1.699	1.311	.683
Large (z)	2.575	2.327	1.960	1.645	1.282	.675

Small Sample Test: Paired t-test

When to use ?

If you want to check, effectiveness of a new treatment, a new method employed or a training one can apply paired t-test. Here observations on every unit are made before and after applying the treatment or method. Hence if treatment or method is effective, there will be significant difference in observations before applying it and after applying it.

$$t = \frac{\overline{x}_D - \mu_0}{s_D / \sqrt{n}}$$

Where,

$\overline{x}_D$ = Mean of differences of 'n' paired observations (samples).

S_D = S.D. of differences of 'n' paired observations (samples).

μ_0 = Mean of differences of 'n' paired observations under $H_{0\ (samples)}$.

Hypothesis on : Paired t-test

Here we test whether the given two sample belongs to same population or not. This can also be taken as there is no significant difference in samples before and after applying some treatment or method.

- H_0: Mean of samples before and after treatment is same.
- H_a: Mean of samples before and after treatment is not same.

Small Sample Test: Chi-square test

When to use ?

If you have attribute data and if you want to test whether any two attributes are associated with each other or they are independent of each other, we can use Chi-Square test for checking independence of attributes.

$$\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i}$$

Where,

O_i = Observed frequency or Cell count

E_i = Expected frequency or Cell count

Hypothesis on : Chi-Square test

Here we test whether attributes in the given data are independent of each other or not.

- H_0: Two attributes are independent of each other.
- H_a: Two attributes are not independent of each other.

Small Sample Test: F test

When to use ?

If you have two different samples and you want to know whether these two sample have same population variance or not, one can use F test.

$$F = \frac{s_1^2}{s_2^2}$$

Where,

S_1^2 = Sample variance of first sample of size n_1

S_2^2 = Sample variance of second sample of size n_2

Hypothesis on : F test

Here we test whether two samples have same variance or not.

- H_0: Variance ratio is one
- H_a: Variance ratio not equal to one