Lecture - 05

→ Hypothesis testing (Type 1 and Type 2 error)

→ One tailed and Two tailed test.

→ Confidence Interval

→ Z-test, T-test, Chi-Square Test

Type 1 and Type 2 error:

Type 1 Error:

Rejecting a null hypothesis Ho, when Ho is true.

Type 2 Error:

Accepting a null hypothesis Ho, when Ho is false.

	Null Hypothesis, Ho	
	Ho is True	Ho is false
Accept Ho	Correct decision	Type II Error → Prob - β
Reject Ho	Type I Error ↓ Prob - α	Correct decision

For above Coin Example,

Null Hypothesis (H_0) = Coin is fair.

Alternate Hypothesis (H_1) = Coin is not fair.

<u>Consider as reality</u>:

Null hypothesis is True or Null hypothesis is false.

<u>Decision made</u>:

Null hypothesis is true or Null hypothesis is false.

<u>Outcomes</u>: [written based on tables]

Outcome 1: We reject the null hypothesis, when in reality it is false. [Correct decision]

Outcome 2: We reject the null hypothesis, when in reality it is true. [Type I Error]

Outcome 3: We accept the null hypothesis, when in reality it is false. [Type II Error]

Outcome 4: We accept the null hypothesis, when in reality it is True. [Correct decision].

In real world Scenario, This hypothesis testing type I and type II error fed into one matrix. It is nothing but Confusion Matrix.

	P	N
T	TP	TN
F	FP	FN

→ Type 2 error

Type 1 error

One Tail and Two Tail Test:

The Usage of one tail and Two tail test is based on the Problem statement.

Example:

Colleges in karnataka have an 85% placement rate. A new college was recently opened and it was found that a Sample of 150 students had a placement rate of 88% with Standard deviation 4%. Does this College has a different placement rate?

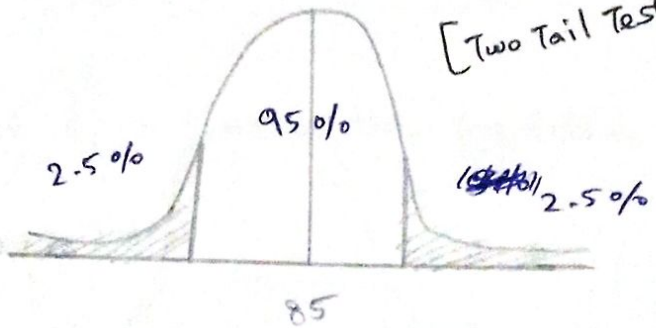
Note that Question, Does this college has a <u>different</u> <u>Placement rate</u>?

The placement rate may be lower than 85%

(or) higher than 85%. So, we use Two-Tail Test.

$\alpha = 0.05$

[Two Tail Test]



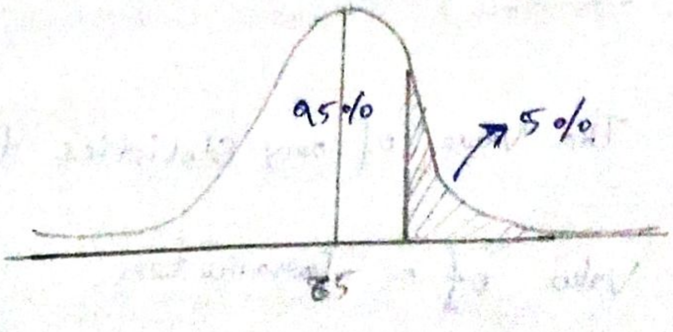
Suppose, the Question is does this College has above 85% of placement rate?

In this Case, placement rate only above 85%. So

Obiviously use One-tail test.

$\alpha = 0.05$

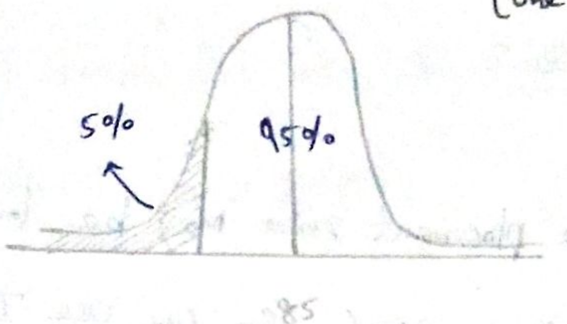
[one tail test]



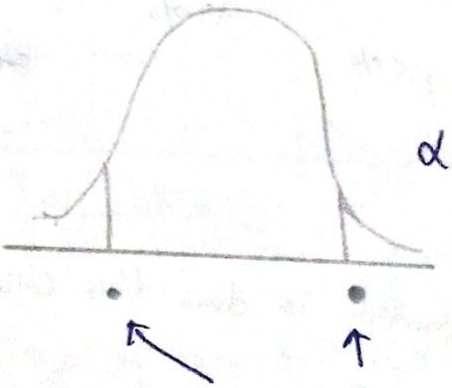
Suppose, the Question is does this college has below 85% of placement rate?

$\alpha = 0.05$

[one - tail test]



Confidence Interval: [Respect to mean]



$\alpha = 0.05$

Find out this Critical Value

So, we know about Point Estimate.

Point Estimate:

The Value of any Statistics that estimates the Value of a Parameter.

In Inferential Statistics, The main work is to be,

1. We need to estimate the population from the Sample.

$$\text{Sample} \xrightarrow{\text{Estimate}} \text{Population}$$

2. We need to estimate the population Mean from the Sample mean.

$$\text{Sample} \xrightarrow{\text{estimate}} \text{Population}$$

Mean   Mean

$$\bar{x} \xrightarrow{\text{estimate}} \mu$$

Remember, The Sample Mean is approximately equal to Population Mean. It may be low, high or may be equal.

Suppose, $\bar{x} = 2.9$ and μ may be 3.

<u>Margin of Error:</u>

The Value of estimation should be some different between the Original mean. We Say it is called as Margin of Error.

Confidence Interval,

Point Estimate ± Marginal error (or) Margin of Error.

Example, (Z-test) problem

1. On the Quant test of CAT exam, the standard deviation is known to be 100. A sample of 30 test takers has a mean of 520 score. Construct a 95% CI about the mean?

$\sigma = 100$ $\quad n=30$ $\quad \bar{x} = 520$ $\quad \alpha = 0.05$

The population Standard deviation is known but the population mean does not known.

1. population std is given.
2. $n \geq 30$

$\} \longrightarrow$ Z-test

formula,

Point Estimate $\pm$ Margin of Error.

$$\boxed{\bar{x} \pm Z_{\alpha/2} \boxed{\frac{\sigma}{\sqrt{n}}}} \rightarrow \text{Standard Error}$$

See directly $\rightarrow$ -1.96

Apply one tail

$\uparrow$

$$Z_{\frac{0.05}{2}} = Z_{0.025}$$

$\rightarrow 1 - 0.025$

$= 0.975$

$\hookrightarrow$ see the value in Z-table (0.975)

$\Downarrow$ The value

1.96

$$\text{Upperbound} = \bar{x} + Z_{\frac{0.05}{2}} \frac{100}{\sqrt{30}}$$

$$\text{Lowerbound} = \bar{x} - Z_{\frac{0.05}{2}} \frac{100}{\sqrt{30}}$$

$$\text{Upperbound} = 520 + 1.96\left(100/5.477225575051661\right)$$

$$= 520 + 1.96\left(18.2574185835054\right)$$

$$= 520 + 35.7845404236 7085$$

$$= 555.784540$$

$$\text{Lowerbound} = 520 - 35.78454042367085$$

$$= 484.215459$$

Apply Z value directly,

$$\text{Upperbound} = \bar{x} + Z_{\frac{0.05}{2}} \frac{100}{\sqrt{30}}$$

$$\text{Upperbound} = 520 + (-1.96)(18.2574185835055\not{4})$$

$$= 520 + (-35.7845042367085)$$

$$= 520 - 35.7845042367085$$

$$= 484.215459 \text{ . [low value]}$$

$$\text{lowerbound} = 520 - (-35.7845042367085)$$

$$= 520 + 35.7845042367085$$

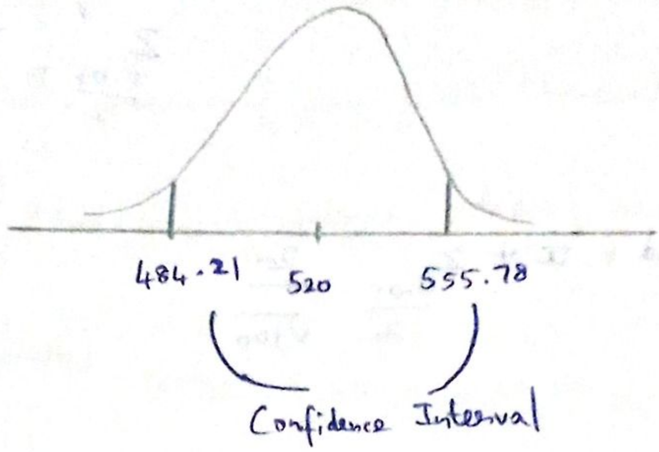
$$= 555.784540 \text{ . [High Value]}$$

So, obiviously we say vice versa,

Upperbound = 555.784540

lowerbound = 484.215459.

Confidence Interval Using Z-test,



Interview Question:

Find the average of the sharks throughout the world?

Soln:

We need to Consider Something,

$\sigma = 20$, $n = 100$, $\bar{x} = 200$ (Size) , $\alpha = 0.05$

↓

Size of Sample Mean

Obivously, we know

$sd = 20$

$n = 100 \geq 30$ } Apply z-test

Formula,

Point Estimate ± Margin of Error

$$\boxed{\bar{x} \pm Z_{\alpha/2} \frac{\sigma}{\sqrt{n}}}$$

This place we apply one tail

$$Z_{\frac{0.05}{2}} = Z_{0.025}$$

$$1 - 0.025$$

$$= 0.975$$

Value

1.96

$$\text{Upperbound} = \bar{x} + Z_{\frac{0.05}{2}} \frac{20}{\sqrt{100}}$$

$$= 200 + (1.96)\left(\frac{20}{10}\right)$$

$$= 200 + 1.96(2)$$

$$= 200 + 3.92$$

$$= 203.92$$

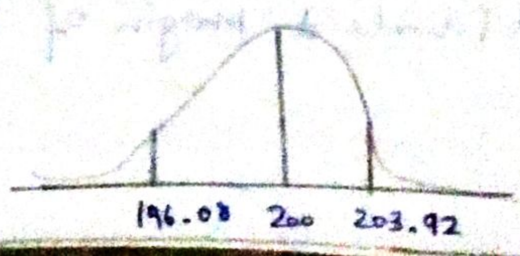
$$\text{Lowerbound} = 200 - (1.96)\left(\frac{20}{10}\right)$$

$$= 200 - 1.96(2)$$

$$= 200 - 3.92$$

$$= 196.08$$

The average size of the Shark between,



(T-Test problem)

1. On the Quant test of CAT exam, a Sample of 30 test takers has a mean of 520 with Standard deviation of 80. Construct 95% of Confidence interval about the mean?

Soln: Given, $n=30$, $\bar{x}=250$, $s=80$, $\alpha=0.05$

Condition:

1. Here population sd is not given.
2. Sample Size $n \geq 30$

$\Big\} \rightarrow$ T-Test

formula for T-Test,

Point Estimate ± Margin of Error

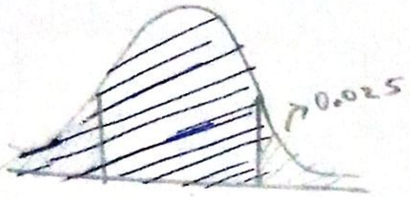
$$\boxed{\bar{x} \pm t_{\alpha/2} \boxed{\left(\frac{s}{\sqrt{n}}\right)}} \rightarrow \text{Standard Error}$$

$$\text{Upper bound} = \bar{x} + t_{\frac{0.05}{2}}\left(\frac{s}{\sqrt{n}}\right)$$

To need to find degree of freedom,

Formula,

$$\text{Degree of freedom} = n-1 = 30-1 = 29$$



$t_{\frac{0.05}{2}} \Rightarrow t_{0.025}$

This place we apply one tail

Two tail $\xrightarrow{\text{Convert}}$ One tail

$1 - 0.025 = 0.975$

Upperbound $= 250 + 2.045 \left(\frac{80}{\sqrt{30}}\right)$

$1 - 0.025$

$= 0.975$

respective value with $df = 29$

$= 250 + 2.045 (14.605934866)$

2.045

$= 250 + 29.869136$

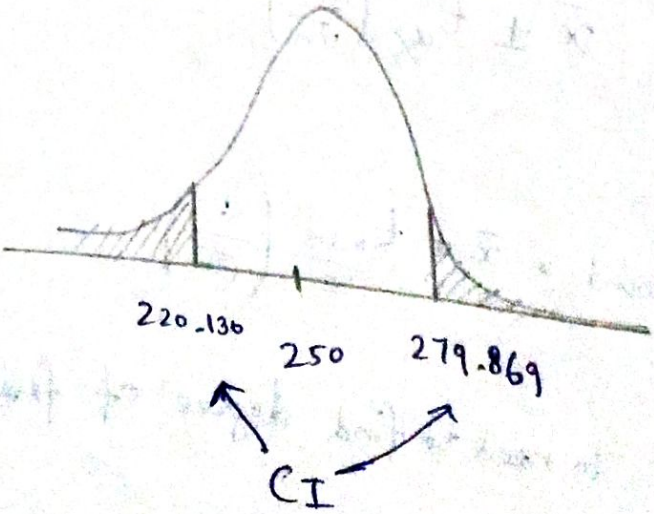
$= 279.86913$

Lowerbound $= 250 - 29.869136$

$= 220.130864$

$220.130 \longleftrightarrow 279.869$

CI



One Sample Z-test:

Conditions:

1. Population std is given.

2. Sample Size $n \geq 30$.

Problem:

1. In the population, The average IQ is 100 with std of 15. Researchers wants to test a new medication to see if there is positive or negative effect on intelligence, or no effect at all. A sample of 30 participants who have taken the medication has a mean of 140. Did the medication affect the intelligence?

Soln: Hypothesis testing (z-test).

1] Null hypothesis

$$H_0 = \mu = 100$$

2] Alternate Hypothesis

$$H_1 = \mu \neq 100$$

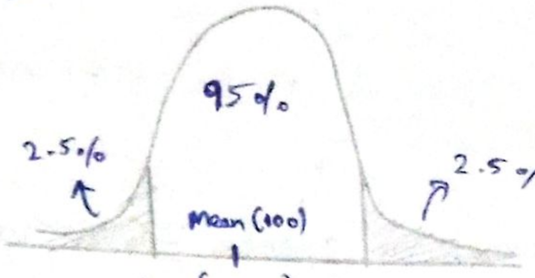
3] State Alpha , $\alpha = 0.05$

4] State Decision rule

$\alpha = 0.05$

whether meditation affect intelligence or not it may increase (or) decrease. So, Two tail Test is used.

2 Tail Test



In Z table, we see the $1 - 0.025 = 0.975$

Corresponding Z-value $\Downarrow$

1.96

5] Calculate Test statistics

$$Z = \frac{\bar{x} - \mu}{\left[\frac{\sigma}{\sqrt{n}}\right]}$$

Standard Error $\rightarrow$ $\left[\frac{\sigma}{\sqrt{n}}\right]$

$Z_{score} = \frac{x_i - \mu}{\sigma}$ [This for when Sample size is 1]

More Sample Size, so $\sqrt{n}$ is used

$$Z = \frac{140 - 100}{\frac{15}{\sqrt{30}}}$$

$$= \frac{40 \times \sqrt{30}}{15}$$

$$\boxed{Z = 14.60}$$

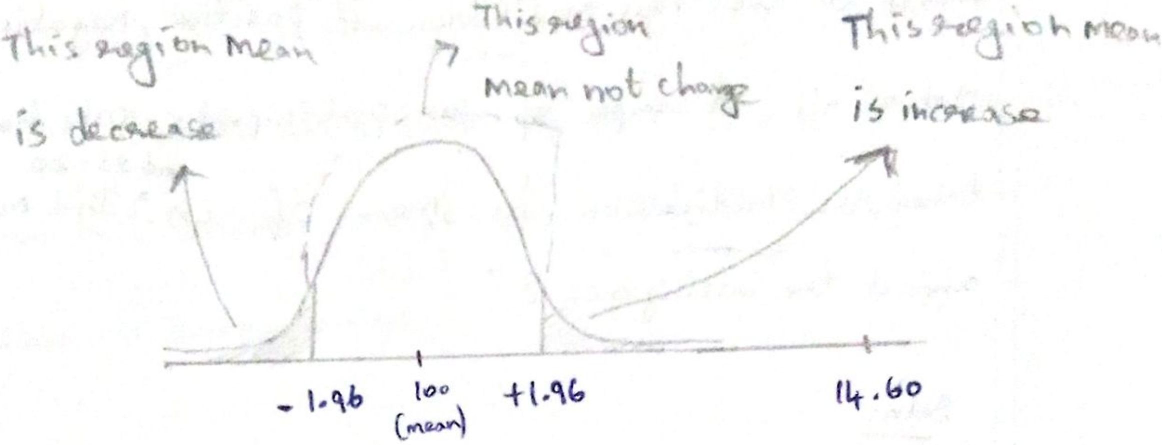
x_i for data points known.

$\bar{x}$ for Sample mean is known

6] State Our Decision

14.60 > 1.96

If Z is less than - 1.96 (or) greater than 1.96, reject the null the hypothesis.



So, we reject the null hypothesis, Obiviously the Population Mean is not equal to 100.

There is two possibility one is mean is decreasing (or) increasing.

According to test statistics, The Value is greater than + 1.96. So, the mean far away from the 14.60.

Mean Value ↑ increase, So the IQ level ↑

Therefore, The meditation Should be Positive for an intelligence.

One Sample T-Test:

t-test ⇒ Unknown Population Std.

1. In the population, The average IQ is 100. Researchers wants to test new meditation if positive, negative (or) not at all. A sample of 30 participants who have taken the medication has mean of 140 and sd = 20. Did medication affect the intelligence?

Soln:

$\mu = 100, \quad n = 30, \quad \bar{x} = 140, \quad sd = 20$

Hypothesis Testing (T-Test)

1] Null Hypothesis, (H_0)

$\mu = 100$

2] Alternate Hypothesis, (H_1)

$\mu \neq 100$

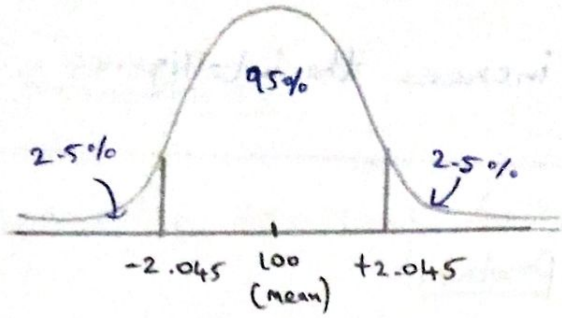
3] State Significance Value $\alpha = 0.05$

4] Degree of freedom $(n-1) = 30-1 = 29$

5] State Decision rule

$\boxed{\alpha = 0.05}$

$1 - 0.025 = 0.975$



$t_{0.975}$

$\Downarrow$ corresponding value with df

2.045

(or) directly see t table of 0.5, $t_{0.5}$ ~~because~~ in two tail because $0.025 + 0.025 = 0.5$. So, The Value is Same 2.045.

6] Calculate Test Statistics:

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

$$t = \frac{140 - 100}{\frac{20}{\sqrt{30}}} \Rightarrow t = \frac{\cancel{40}^{2}}{\cancel{20}} \times \sqrt{30}$$

$$t = 2 \times \sqrt{30} \Rightarrow \boxed{t = 10.95}$$

7] State Our decision:

$10.95 > 2.045$

So, Reject the null hypothesis.

So, Pvalue $\leq$ Significance value

Obiviously Mean value increase, Therefore the Meditation increase the intelligence.

Real world Problem:

$\boxed{\text{Bank}} \xrightarrow[\text{open}]{\text{wants to}} \boxed{\text{ATM Machine On Area}}$

Shencottai Population - 2023 [36,000]

percentage of ATM used - 40%

Mean (μ) = 14400

$n = 10000$

$\bar{x} = 4000$

Population Standard deviation (σ) = 1000

Hypothesis testing (Z-test)

(i) Null hypothesis $(H_0) = \mu = 14400$

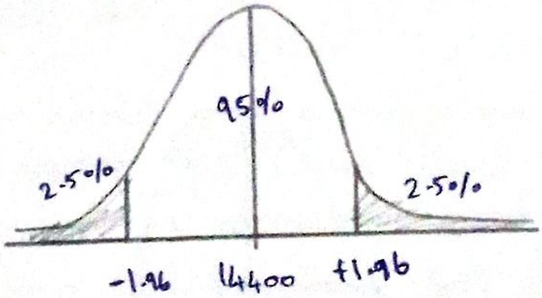
(ii) Alternate Hypothesis $(H_1) = \mu \neq 14400$

(iii) State Alpha, $\alpha = 0.05$

(iv) State Decision rule,

Wheather the ATM wants to open or not.

2 Tail Test, $\alpha = 0.05$



2.5% are both are Symmetric. So, find any one side,

$1 - 0.025 = 0.975$

Correspond Z-value

1.96

(v) Calculate Test stats:

$$Z = \frac{\bar{x} - \mu}{\left[\frac{\sigma}{\sqrt{n}}\right]}$$

→ Standard Error

$$Z = \left[\frac{4000 - 14400}{\frac{1000}{\sqrt{10000}}}\right]$$

$$Z = \frac{-10,400}{1000} \times 100$$

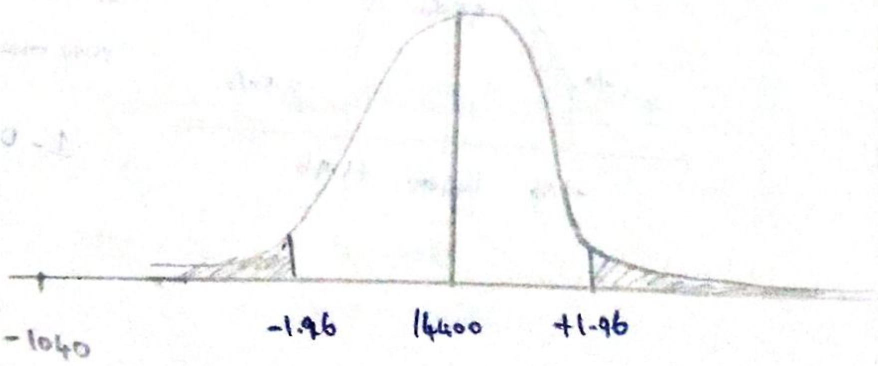
$$\boxed{Z = -1040}$$

(vi) State decision;

$-1040 < -1.96$

If z is less than -1.96, So reject the null hypothesis.

So, we reject the null hypothesis, obiviously the population Mean is not equal to ~~14~~ 14400.



There is two possibility one is Mean decrease (or) increase.

The -1040 is lesser than -1.96.

Mean Value ↓ decrease, so obivously the percentage of ATM used people also ↓ decrease. So, ATM is not opened is a better Choice.